Bidimensional Random Effect Estimation in Mixed Stochastic Differential Model

Charlotte Dion, Valentine Genon-Catalot

▶ To cite this version:

Charlotte Dion, Valentine Genon-Catalot. Bidimensional Random Effect Estimation in Mixed Stochastic Differential Model. Statistical Inference for Stochastic Processes, 2016, 19 (2), pp.131-158. 10.1007/s11203-015-9122-0 . hal-01103303v2

HAL Id: hal-01103303
https://hal.science/hal-01103303v2

Submitted on 22 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

BIDIMENSIONAL RANDOM EFFECT ESTIMATION IN MIXED STOCHASTIC DIFFERENTIAL MODEL

C. DION [1,2] AND V. GENON-CATALOT [2]

(1)*LJK, UMR CNRS 5224, Université Joseph Fourier,*
51 rue des Mathématiques, 38041 Grenoble

(2)*MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité,*
45 rue des Saints Pères, 75006 Paris

Abstract

In this work, a mixed stochastic differential model is studied with two random effects in the drift. We assume that N trajectories are continuously observed throughout a large time interval $[0, T]$. Two directions are investigated. First we estimate the random effects from one trajectory and give a bound of the L^2-risk of the estimators. Secondly, we build a nonparametric estimator of the common bivariate density of the random effects. The mean integrated squared error is studied. The performances of the density estimator are illustrated on simulations.

AMS Subject Classification 62M05, 62G07, 60J60

Keywords Adaptive bandwidth, density estimation, kernel estimator, mean integrated squared error, mixed-effects models, stochastic differential equations.

1. Introduction

Mixed-effects models are used to analyse repeated measurements with similar functional form but with some variability between experiments (see *e.g.* Davidian and Giltinan, 1995; Pinheiro and Bates, 2000; Diggle *et al.*, 2002). The advantage is that a single estimation procedure is used to fit the overall data simultaneously. In the present work the model of interest is a mixed-effects stochastic differential equation (SDE). Each equation represents the behaviour of one subject and corresponds to one realization of the random effects. Hence the random effects represent the particularity of each process. Mixed-effects SDEs have various applications such as neuronal or pharmacokinetic modelling (see *e.g.* Picchini *et al.*, 2008; Donnet and Samson, 2013).

Estimation methods in SDEs with random effects have been proposed in literature. The main purposes are the estimation of the common distribution of the random effects in a parametric or nonparametric way. The estimation of the common density of the random effects is mainly parametric. Most methods assume normality of the random effects and estimate the population parameters (see *e.g.* Picchini *et al.*, 2008; Delattre *et al.*, 2013; Delattre and Lavielle, 2013). However, one can ask if this assumption is reasonable in some application contexts. Nonparametric estimation can allow us to get around this problem. To the best of our knowledge, the first reference in this context is Comte *et al.* (2013) who propose different nonparametric estimators and then Dion (2014) who develops two adaptive nonparametric estimators for the Orsntein-Uhlenbeck model with an application to a neuronal database. But these two references focus on a one-dimensional random effect.

In the present work we study the case of two random effects or, in other words, of one bidimensional random effect. We want to deal with two points: the estimation of the random effects and the nonparametric estimation of their common density. This bivariate context makes the study more complex. In fact the estimation of the random effects is done using matrix norm and operator, and anisotropy appears in the density estimation part. We consider N trajectories, observed on the interval $[0,T]$ where T is given. For $j=1,...,N$ the dynamics of each process is described by the stochastic differential equation

$$\begin{cases} dX_j(t) &= b^t(X_j(t))\phi_j dt + \sigma(X_j(t))dW_j(t) \\ X_j(0) &= \gamma_j \end{cases} \tag{1}$$

where $\phi_j = (\phi_{j,1}, \phi_{j,2})^t \in \mathbb{R}^2$ is the bidimensional random effect, $b(.) = (b_1(.), b_2(.))^t$, $\sigma(.)$ are known functions defined on $\mathbb{R}$, $(W_j)_{1\leq j\leq N}$ are N independent Wiener processes and γ_j is a real valued random variable. The random variables $((\phi_1,\gamma_1),\dots,(\phi_N,\gamma_N))$ are *i.i.d.* and the sequences $((\phi_1,\gamma_1),\dots,(\phi_N,\gamma_N))$ and $(W_1,\dots,W_N)$ are independent. The two random effects are not assumed independent. This model is more general than the ones investigated in Comte *et al.* (2013) where the drift has the form $b(.)\phi$ or $b(.)+\phi$ with ϕ a real valued random effect. Notice that the N trajectories $(X_j(t), 0\leq t\leq T)_{1\leq j\leq N}$ are *i.i.d.*.
We assume that the $\phi_j = (\phi_{j,1},\phi_{j,2})^t$'s have a common bivariate density f. Our goal is twofold: first estimate the random effects ϕ_j's and then their density f, from the observations $(X_j(t), 0\leq t\leq T)_{1\leq j\leq N}$, with large T and N.
The estimation of the random variables ϕ_j's follows the steps of Genon-Catalot and Larédo (2014) where only one multiplicative random effect in the drift is considered. We build an estimator $\widehat{A}_j(T)$ of ϕ_j based on the trajectory $(X_j(t), 0\leq t\leq T)$ and study its L^2-risk. This leads to a bound of order $1/T$. Then we propose a kernel estimator of the density f which uses the sample $(\widehat{A}_j(T))_j$. When f is in a Nikol'ski space a bound of the mean integrated squared error is established and the rate of convergence is evaluated. Finally a data-driven choice of the bandwidth based on a Goldenshluger and Lepski's criterion for anisotropic multi-index is proposed (see *e.g.* Kerkyacharian *et al.*, 2007; Goldenshluger and Lepski, 2011) and leads to an adaptive estimator.
Section 2 is dedicated to assumptions and definitions of some useful quantities for the estimation of ϕ_j. In Section 3 the estimator of the random effects is built and its L^2-risk is bounded. In particular we deal with two main examples: the Ornstein-Uhlenbeck model and the Cox-Ingersoll-Ross model. In Section 4 the estimator of the density f is studied. Finally, Section 6 is devoted to numerical simulations to illustrate estimators. Proofs are relegated in Section 8.

2. Notation and assumptions

2.1. **General assumptions on the model.** Consider real valued processes $(X_j(t))_{j=1,\dots,N}$ given by (1). We assume that $(W_j)_{j=1\dots,N}$ and $(\phi_j,\gamma_j)_{j=1,\dots,N}$ are defined on a common probability space $(\Omega,\mathcal{F},\mathbb{P})$. Consider the following assumptions (see *e.g.* Kutoyants, 2004; Kessler *et al.*, 2012, for details).

(A1) The real valued functions $x\mapsto b_1(x)$, $x\mapsto b_2(x)$ and $x\mapsto\sigma(x)$ are continuous on $\mathbb{R}$ and b_1, b_2 Lipschitz and σ Hölder with exponent belonging to $[1/2,1]$.

(A2) There exists an open set Φ of $\mathbb{R}^2$ and an interval $(l,r)\subset\mathbb{R}$ such that $\sigma^2(x)>0$ for $x\in(l,r)$, and for all $\varphi\in\Phi$ the function $s_\varphi : x\mapsto \exp(-2\int_{x_0}^x \frac{b^t(u)\varphi}{\sigma^2(u)}du)$, $x_0\in(l,r)$ satisfies

$$\int_l s_\varphi(y)dy = +\infty, \quad \int^r s_\varphi(y)dy = +\infty.$$

The function $m_\varphi : x\mapsto 1/(\sigma^2(x)s_\varphi(x))$ satisfies $M(\varphi)=\int_l^r m_\varphi(x)dx < +\infty$. We set

$$\pi_\varphi(x) := \mathbb{1}_{(l,r)}(x)\frac{m_\varphi(x)}{M(\varphi)}. \tag{2}$$

(A3) ϕ_j takes values in Φ, has distribution $f(\varphi)d\varphi$ and (ϕ_j,γ_j) has distribution on $\Phi\times(l,r)$

$$\pi(d\varphi,dx)=f(\varphi)d\varphi\otimes\pi_\varphi(x)dx.$$

Assumption **(A1)** ensures the existence and uniqueness of a strong solution $X_j(.)$ to (1) for all random variables $(\phi_j,\gamma_j)\in\mathbb{R}^3$, adapted to the filtration $(\mathcal{F}_t=\sigma((\phi_j,\gamma_j)_{j=1,\ldots,N},\ W_j(s),s\le t,\ j=1\ldots,N),t\ge 0)$. Then (ϕ_j,γ_j) is $\mathcal{F}_0$-measurable and W_j is a $(\mathcal{F}_t,t\ge 0)$ Brownian motion.

In the following, for $\varphi=(\varphi_1,\varphi_2)^t$ a fixed value in Φ we denote by X^φ the process solution of the stochastic differential equation with fixed φ:

$$\begin{cases} dX_j^\varphi(t)= & b^t(X_j^\varphi(t))\varphi dt+\sigma(X_j^\varphi(t))dW_j(t)\\ X_j^\varphi(0)\sim & \pi_\varphi,\ X_j^\varphi(0)\text{ independent of }W_j.\end{cases} \tag{3}$$

Under **(A1)-(A2)**, for $\varphi\in\Phi$, the process defined by (3) is strictly stationary and ergodic with scale density s_φ, speed density m_φ and marginal distribution $\pi_\varphi(x)dx$. Under **(A1)-(A3)**, according to Genon-Catalot and Larédo (2014), the conditional distribution of X_j given $\phi_j=\varphi$ is identical to the distribution of (3), the process $((\phi_j,X_j(t)),t\ge 0)$ is strictly stationary with marginal distribution π and $\mathbb{P}(X_j(t)\in(l,r),\forall t>0)=1$.
Finally under **(A1)-(A3)** as for $j=1,\ldots,N$, for all $t\ge 0$, $X_j(t)\in(l,r)$, $\forall T>0$:

$$\int_0^T\frac{b_1^2}{\sigma^2}(X_j(s))ds<+\infty,\ \int_0^T\frac{b_2^2}{\sigma^2}(X_j(s))ds<+\infty,\quad a.s.$$

2.2. **Key examples.** We investigate two classical examples: the mixed Ornstein-Uhlenbeck model and the mixed Cox-Ingersoll-Ross model.

Example. *[1] The Ornstein Uhlenbeck model (O-U) with two random effects is defined as*

$$\begin{cases} dX_j(t) &=(\phi_{j,1}-\phi_{j,2}X_j(t))dt+\sigma dW_j(t)\\ X_j(0) &=\gamma_j\end{cases} \tag{4}$$

with $\sigma(x)=\sigma>0$, $(l,r)=\mathbb{R}$, *and* $b(x)=(1,-x)^t$. *Assumption* ***(A2)*** *requires* $\Phi=\mathbb{R}\times(0,+\infty)$ *and leads to* $\pi_\varphi=\mathcal{N}\left(\varphi_1/\varphi_2,\sigma^2/(2\varphi_2)\right)$ *as invariant distribution for fixed* φ.

Example. *[2] The Cox-Ingersoll-Ross model (C-I-R) with two random effects is defined as*

$$\begin{cases} dX_j(t) &=(\phi_{j,1}-\phi_{j,2}X_j(t))dt+\sigma\sqrt{X_j(t)}dW_j(t)\\ X_j(0) &=\gamma_j\end{cases} \tag{5}$$

with $\sigma(x)=\sigma\sqrt{x}$ *with* $\sigma>0$, *and* $b(x)=(1,-x)^t$. *Assumption* ***(A2)*** *requires* $2\varphi_1/\sigma^2\ge 1$ *and* $2\varphi_2/\sigma^2>0$ *(in particular the process is always positive). This leads to consider* $\Phi=(\sigma^2/2,+\infty)\times(0,+\infty)$. *The invariant distribution, for fixed* φ, $\pi_\varphi=\Gamma(2\varphi_1/\sigma^2,\sigma^2/(2\varphi_2))$ *is the Gamma distribution with shape parameter* $2\varphi_1/\sigma^2$ *and scale parameter* $\sigma^2/(2\varphi_2)$.

2.3. **Specific assumptions and notations for estimation.** Our estimation is based on the following quantities. As in Comte *et al.* (2013) we define for $j=1,...,N$,

$$U_j(T):=\int_0^T\frac{b}{\sigma^2}(X_j(s))dX_j(s), \tag{6}$$

which is a column vector with size 2×1 and the 2×2 symmetric matrix:

$$V_j(T):=\int_0^T\frac{b\,b^t}{\sigma^2}(X_j(s))ds=\begin{pmatrix}\int_0^T\frac{b_1^2}{\sigma^2}(X_j(s))ds & \int_0^T\frac{b_1b_2}{\sigma^2}(X_j(s))ds\\ \int_0^T\frac{b_1b_2}{\sigma^2}(X_j(s))ds & \int_0^T\frac{b_2^2}{\sigma^2}(X_j(s))ds\end{pmatrix}. \tag{7}$$

Using (1), $U_j(T)$ can be decomposed as follows

$$U_j(T)=V_j(T)\phi_j+M_j(T)\text{ with }M_j(T):=\int_0^T\frac{b}{\sigma}(X_j(s))dW_j(s). \tag{8}$$

Note that

$$V_j(T) =< M_j >_T \tag{9}$$

where $< . >$ is the quadratic variation of a continuous local martingale. For all measurable function $h : (l,r) \to \mathbb{R}$ π_φ-integrable for all $\varphi \in \Phi$, we define the random variable $\pi_{\phi_j}(h) := \int_l^r h(x)\pi_{\phi_j}(x)dx$.

(A4) For $\varphi \in \Phi$ and $i \in \{1,2\}$, we assume

$$\pi_\varphi\left(\frac{b_i^2}{\sigma^2}\right) = \int_l^r \frac{b_i^2}{\sigma^2}(x)\pi_\varphi(x)dx < +\infty.$$

We define the random matrix

$$L_j := \pi_{\phi_j}\left(\frac{b\,b^t}{\sigma^2}\right).$$

(A5) For $j = 1,\dots,N$, $\mathbb{P}\left(L_j \text{ invertible}\right) = 1$.

Under **(A1)-(A4)**, Theorem 3.1. of Genon-Catalot and Larédo (2014) gives:

$$\frac{< M_j >_t}{t} = \frac{V_j(t)}{t} \underset{t\to+\infty}{\to} L_j,\ a.s. \tag{10}$$

One can notice that assumption **(A5)** implies that $V_j(t)$ is invertible for t large enough.
In the study we denote $\|.\|$ the $L^2(\mathbb{R})$-norm, $\|.\|_p$ the $L^p(\mathbb{R})$-norm when $p \neq 2$, $\|\|_2$ the euclidean norm of $\mathbb{R}^2$ and $\|\|_F$ the Frobenius norm of matrices defined for $A \in M_2(\mathbb{R})$ by $\|A\|_F^2 = \sum_{i,j} A_{i,j}^2 = \mathrm{Tr}(\,A^tA)$. We denote $S_2(\mathbb{R})$ the subset of symmetric matrices of $M_2(\mathbb{R})$.

3. Random effect estimation

We define first for each j an estimator of the random variable ϕ_j based on the trajectory $X_j(t),\ t \in [0,T]$. Then these estimators are used to build an estimator of the density f.

3.1. Definition of the estimator of the random effects.

Let us define N bidimensional random variables

$$A_j(T) := \left(A_{j,1}(T), A_{j,2}(T)\right)^t = V_j(T)^{-1}U_j(T), \tag{11}$$

which corresponds to the maximum likelihood estimator of ϕ_j when $\phi_j = \varphi$ is deterministic. Note that

$$A_j(T) = \phi_j + V_j(T)^{-1}M_j(T). \tag{12}$$

Thus $A_j(T)$ is a consistent estimator of ϕ_j as T tends to infinity according to convergence (10). Because of the presence of the inverse matrix $V_j(T)^{-1}$ in formulae (11)-(12), it is difficult to prove that $A_j(T)$ has finite moments and to compute any of them. To overcome this theoretical difficulty, we consider a truncated estimator of ϕ_j. Let

$$\widehat{A_j}(T) := A_j(T)\mathbb{1}_{B_j(T)},\quad B_j(T) := \{V_j(T) \geq \kappa\sqrt{T}I_2\} = \{\min(\lambda_{1,j}(T), \lambda_{2,j}(T)) \geq \kappa\sqrt{T}\} \tag{13}$$

where $\lambda_{i,j}(T),\ i = 1,2$ are the eigenvalues of $V_j(T)$ and I_2 is the identity matrix of $M_2(\mathbb{R})$. The inequality in the definition of $B_j(T)$ has a matrix sense: for two matrices $(A,B) \in S_2(\mathbb{R})$, $A \leq B$ if and only if $B - A$ is a non negative matrix (see Appendix 8.5).

Relations (7) and (9) show that $V_j(T)$ is a non negative symmetric matrix, thus its eigenvalues are non negative. We are able to bound the L^2-risk of the estimator $\widehat{A}_j(T)$ of ϕ_j. This bound is needed to evaluate the mean integrated squared error of the nonparametric estimator of the density f.

3.2. **Main result.** We denote $\mathcal{L}_\varphi$ the infinitesimal generator of the process (3), given for $F \in \mathcal{C}^2((l,r))$, by $\mathcal{L}_\varphi F(x) := (\sigma^2(x)/2)F''(x) + (\,b^t(x)\varphi)F'(x)$. Its domain is included in $L^2_{\pi_\varphi}$ which is the space of functions f such that $\int_l^r f^2 d\pi_\varphi < \infty$ (for details see *e.g.* Genon-Catalot *et al.* (2000)). When F is a matrix, the notation $\mathcal{L}_\varphi F$ indicates that we apply the operator on each coefficient of the matrix. We also define analogously F' as the matrix of derivatives. For all $g = (g_{i,k})_{1\le i,k\le 2} \in M_2(\mathbb{R})$ such that $\pi_\varphi(g_{i,k}^2) < +\infty$ for all $\varphi \in \Phi$, we associate the matrix

$$F_\varphi^g = \begin{pmatrix} F_\varphi^{g_{1,1}} & F_\varphi^{g_{1,2}} \\ F_\varphi^{g_{2,1}} & F_\varphi^{g_{2,2}} \end{pmatrix}$$

satisfying $-\mathcal{L}_\varphi F_\varphi^g = g - \pi_\varphi g$ for all $\varphi \in \Phi$. We denote for simplicity

$$H_{\phi_j} := F_{\phi_j}^{b\,b^t/\sigma^2}. \tag{14}$$

Examples below show how H_{ϕ_j} can be constructed. We are now able to announce the main result on the estimator on the random effects.

Proposition 1. *Consider the processes $(X_j(t), j = 1,\dots,N)$ given by (1) under **(A1)-(A5)**. Assume that for $j = 1,\dots,N$,*

$$\mathbb{E}\left[\frac{\|L_j\|_F^2}{[\det(L_j)]^2}\mathrm{Tr}\,(L_j)\right] < \infty, \quad \mathbb{E}\left[\frac{\|L_j\|_F^4}{[\det(L_j)]^4}\mathrm{Tr}\left(\pi_{\phi_j}\left(\left(\frac{b\,b^t}{\sigma^2}\right)^2\right)\right)\right] < \infty,$$

$$\mathbb{E}\left[\|\phi_j\|_2^2\left(\frac{1}{L_{j,1,1}^2} + \frac{1}{L_{j,2,2}^2}\right)\pi_{\phi_j}(\|H_{\phi_j}\|_F^2)\right] + \mathbb{E}\left[\|\phi_j\|_2^2\left(\frac{1}{L_{j,1,1}^2} + \frac{1}{L_{j,2,2}^2}\right)\pi_{\phi_j}(\|H'_{\phi_j}\|_F^2\sigma^2)\right] < \infty$$

and

$$\mathbb{E}\left[\pi_{\phi_j}(\|H_{\phi_j}\|_F^4)\right] + \mathbb{E}\left[\pi_{\phi_j}(\|H'_{\phi_j}\|_F^4\sigma^4)\right] < \infty.$$

Then, there exists a constant $C > 0$ such that

$$\mathbb{E}\left[\|\widehat{A}_j(T) - \phi_j\|_2^2\right] \le \frac{C}{T}. \tag{15}$$

The two last assumptions of Proposition 1 correspond to the application of Proposition 7 given in Section 8.1 with $g = b\,b^t/\sigma^2$, $p = 1$ and $p = 2$. This proposition is the key of the nonparametric estimation procedure set up in Section 4.

3.3. **Key examples continued.** Let us investigate the two examples given in Section 2.2.

Example. ***[1]****(continued) In this case*

$$\frac{b\,b^t}{\sigma^2}(x) = \frac{1}{\sigma^2}\begin{pmatrix} 1 & -x \\ -x & x^2 \end{pmatrix}, \quad V_j(T) = \frac{1}{\sigma^2}\begin{pmatrix} T & -\int_0^T X_j(s)ds \\ -\int_0^T X_j(s)ds & \int_0^T X_j(s)^2 ds \end{pmatrix}.$$

For simplicity we set here $\phi_j = \phi = (\phi_1, \phi_2)^t$. The limit matrix L of $V(T)/T$ is given by

$$L = \pi_\phi\left(\frac{b\,b^t}{\sigma^2}\right) = \frac{1}{\sigma^2}\begin{pmatrix} 1 & -\frac{\phi_1}{\phi_2} \\ -\frac{\phi_1}{\phi_2} & \frac{\sigma^2}{2\phi_2} + \frac{\phi_1^2}{\phi_2^2} \end{pmatrix}, \ det(L) = \frac{\sigma^2}{2\phi_2} > 0.$$

We can check that

$$H_\phi(x) = \begin{pmatrix} 1 & -\frac{1}{\phi_2\sigma^2}(x - \frac{\phi_1}{\phi_2}) \\ -\frac{1}{\phi_2\sigma^2}(x - \frac{\phi_1}{\phi_2}) & \frac{2\phi_1}{\sigma^2\phi_2^2}\left(x - \frac{\phi_1}{\phi_2}\right) + \frac{1}{2\sigma^2\phi_2}\left[\left(x - \frac{\phi_1}{\phi_2}\right)^2 - \frac{\sigma^2}{2\phi_2}\right] \end{pmatrix}.$$

Details are in Appendix. The assumptions of Proposition 1 are fulfilled if $\mathbb{E}[\phi_1^6 + \phi_2^{-10}] < \infty$.

Example. *[2](continued) In this case,*

$$\frac{b\,b^t}{\sigma^2}(x) = \frac{1}{\sigma^2}\begin{pmatrix} \frac{1}{x} & -1 \\ -1 & x \end{pmatrix}, \quad V_j(T) = \frac{1}{\sigma^2}\begin{pmatrix} \int_0^T \frac{1}{X_j(s)}ds & -T \\ -T & \int_0^T X_j(s)ds \end{pmatrix}.$$

For $\Phi = (\sigma^2, +\infty) \times (0, +\infty)$, $\pi_\varphi(x \mapsto 1/x^2) < \infty$ *for all* $\varphi \in \Phi$ *and the limit matrix* L *is given by:*

$$L = \pi_\phi\left(\frac{b\,b^t}{\sigma^2}\right) = \frac{1}{\sigma^2}\begin{pmatrix} \frac{2\phi_2}{2\phi_1-\sigma^2} & -1 \\ -1 & \frac{\phi_1}{\phi_2}. \end{pmatrix}, \quad \det(L) = \frac{\sigma^2}{2\phi_1 - \sigma^2} > 0.$$

One can verify (see Appendix) that

$$H_\phi(x) = \begin{pmatrix} \frac{2}{\sigma^2(2\phi_1-\sigma^2)}\log(x) & 1 \\ 1 & \frac{1}{\phi_2\sigma^2}\left(x - \frac{\phi_1}{\phi_2}\right) \end{pmatrix}.$$

The details are relegated in Appendix. For the assumptions of Proposition 1 we must have $\pi_\phi(x \mapsto |\log(x)|)$, $\pi_\phi(x \mapsto 1/x^2)$ *and* $\pi_\phi(x \mapsto 1/x^4)$ *finite. The last condition imposes a reduction of* Φ *to* $\Phi = (2\sigma^2, +\infty) \times (0, +\infty)$. *Moreover the other assumptions are fulfilled if* $\mathbb{E}[\log^2(\phi_2) + \phi_1^5 + \phi_2^8 + \phi_2^{-4} + \psi^2(2\phi_1/\sigma^2 - 1)] < \infty$, *where* $\psi(x) := \Gamma'(x)/\Gamma(x)$ *is the digamma function.*

4. Nonparametric estimation

In this section we assume that $f \in \mathbb{L}^2(\mathbb{R}^2)$ and set up a nonparametric estimation procedure based on a kernel estimator. We study the obtained estimator and determinate its rate of convergence.

4.1. **Nonparametric estimator of the density of the random effects.** By Proposition 1, $\widehat{A}_j(T)$ is a consistent estimator of ϕ_j when T tends to infinity. It is therefore natural to define a kernel estimator based on the $\widehat{A}_j(T)$'s. Let us denote by K a kernel in $\mathcal{C}^1(\mathbb{R}^2)$ such that the partial derivatives $\frac{\partial K}{\partial u}$ and $\frac{\partial K}{\partial v}$ are in $\in L^2(\mathbb{R}^2)$, K is integrable, $\iint K(u,v)dudv = 1$ and $\|K\|^2 = \iint K^2(u,v)dudv < +\infty$. For all $h = (h_1, h_2), h_1 > 0, h_2 > 0$, for all $(u,v) \in \mathbb{R}^2$, we denote

$$K_h(u,v) = \frac{1}{h_1 h_2} K\left(\frac{u}{h_1}, \frac{v}{h_2}\right).$$

For example one can consider the Gaussian kernel $K(u,v) = K_1(u) \times K_1(v)$, with $K_1(u) = (1/\sqrt{2\pi})\exp(-u^2/2)$. We define the estimator of the density f for $x = (x_1, x_2) \in \mathbb{R}^2$ by

$$\widehat{f}_h(x) = \frac{1}{N}\sum_{j=1}^N K_h(x - \widehat{A}_j(T)). \tag{16}$$

Denoting $f_h(x) := K_h \star f(x) = \iint f(y_1,y_2)K_h(x_1 - y_1, x_2 - y_2)dy_1dy_2$, the following result holds.

Proposition 2. *Under (A1)-(A5) and under the assumptions of Proposition 1,*

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \le 2\|f - f_h\|^2 + 2\max\left(\frac{1}{h_1^3 h_2}, \frac{1}{h_1 h_2^3}\right)\left(\left\|\frac{\partial K}{\partial u}\right\|^2 + \left\|\frac{\partial K}{\partial v}\right\|^2\right)\frac{C}{T} + \frac{\|K\|^2}{N h_1 h_2} \tag{17}$$

with C *the constant which appears in Proposition 1.*

This bound comes from the bias-variance decomposition:

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \le 2\|f - f_h\|^2 + 2\|\mathbb{E}[\widehat{f}_h] - f_h\|^2 + \mathbb{E}[\|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|^2].$$

The first term is a bias term due to the approximation of f by f_h, it decreases when h_1, h_2 decrease. The third term is a variance term which increases when h_1, h_2 decrease. Finally, the middle term is an error term due to the approximation of the ϕ_j's by the $\widehat{A}_j(T)$'s also increasing when h_1, h_2 decrease. Note that the orders are consistent with the result of Comte *et al.* (2013), Proposition 1 for the multiplicative model. In fact in the case of a single random effect, multiplicative in the drift, the second term has the order $1/(Th^3)$ and the third term: $1/(Nh)$.

To choose the best h, a compromise must be done between between the bias term and the variance and middle terms.

4.2. **Rates of convergence.** We consider anisotropic Nikol'ski classes of functions which are well fitted to evaluate the order of the bias term (see Goldenshluger and Lepski (2011) and Comte and Lacour (2013) for example).

Definition 3. *A function f is in the Nikol'ski class $\mathcal{N}(\beta, R)$ if $f : \mathbb{R}^2 \to \mathbb{R}$ admits partial derivatives of order $\lfloor \beta_i \rfloor$, $i \in \{1,2\}$ such that, with $y_1 = (y,0), y_2 = (0,y)$, for all $y \in \mathbb{R}$, for $i \in \{1,2\}$*

$$\left(\int \left| \frac{\partial^{\lfloor \beta_i \rfloor} f}{(\partial x_i)^{\lfloor \beta_i \rfloor}}((x_1,x_2)+y_i) - \frac{\partial^{\lfloor \beta_i \rfloor} f}{(\partial x_i)^{\lfloor \beta_i \rfloor}}(x_1,x_2) \right|^2 dx \right)^{1/2} \le R|y|^{\beta_i - \lfloor \beta_i \rfloor},$$

$$\left\| \frac{\partial^{\lfloor \beta_i \rfloor} f}{(\partial x_i)^{\lfloor \beta_i \rfloor}} \right\| \le R$$

(with $\lfloor \beta \rfloor$ denotes the largest integer strictly less than β).

Recall that kernel $K : \mathbb{R} \to \mathbb{R}$ is of order $l \in \mathbb{N}^*$ if for $j = 1, \dots, l$, $\int |x|^j |K(x)| dx < +\infty$ and $\int x^j K(x) dx = 0$ (see Tsybakov (2009)). In this context we are now able to bound the bias term $\|f - f_h\|^2$.

Proposition 4. *If $f \in \mathcal{N}(\beta, R)$ and $K(x) = K(x_1,x_2) = K_1(x_1)K_2(x_2)$ where K_1 is a kernel of order $l_1 := \lfloor \beta_1 \rfloor$, K_2 a kernel of order $l_2 := \lfloor \beta_2 \rfloor$, $\int |K_1(u)u^{\beta_1}| du < \infty$ and $\int |K_2(u)u^{\beta_2}| du < \infty$, then there exist two constants $c_1, c_2 > 0$ such that*

$$\|f - f_h\|^2 \le c_1 h_1^{2\beta_1} + c_2 h_2^{2\beta_2}. \tag{18}$$

In order to minimize the MISE, if we only consider the first and the last term of the right hand side of (17), we have to minimize the following function of two variables:

$$(h_1,h_2) \mapsto \xi(h_1,h_2) = \frac{1}{Nh_1h_2} + h_1^{2\beta_1} + h_2^{2\beta_2}.$$

With $1/\bar{\beta} = 1/(2\beta_1) + 1/(2\beta_2)$, we get the minimizing values

$$h_1^* = O\left(N^{-\frac{1}{2\beta_1}\frac{\bar{\beta}}{\bar{\beta}+1}}\right), \quad h_2^* = O\left(N^{-\frac{1}{2\beta_2}\frac{\bar{\beta}}{\bar{\beta}+1}}\right) \tag{19}$$

and this implies

$$\xi(h_1^*,h_2^*) = O\left(N^{-\frac{4\beta_1\beta_2}{(2\beta_1+1)(2\beta_2+1)-1}}\right) = O\left(N^{-\frac{2\bar{\beta}}{2+2\bar{\beta}}}\right).$$

Therefore, we require (see (17))

$$\frac{1}{T}\max\left(\frac{1}{h_1^{*3}h_2^*}, \frac{1}{h_1^*h_2^{*3}}\right) \le \frac{1}{Th_1^*h_2^*}\left(\frac{1}{h_1^{*2}} + \frac{1}{h_2^{*2}}\right) \le N^{-\frac{4\beta_1\beta_2}{(2\beta_1+1)(2\beta_2+1)-1}}.$$

This holds under the following condition

$$T \ge N^{1+\frac{4\beta_1}{(2\beta_1+1)(2\beta_2+1)-1}} + N^{1+\frac{4\beta_2}{(2\beta_1+1)(2\beta_2+1)-1}}. \tag{20}$$

This condition is implied by $T \ge N^3$, and if $\beta_i > 1/2$ by $T \ge N^2$. We are now able to state the order of the risk, for large N and T.

Corollary 5. *Under (A1)-(A5), under the assumptions of Proposition 4, under conditions (19) and (20), we have*

$$\mathbb{E}[\|\widehat{f}_{h^*} - f\|^2] = O\left(N^{-\frac{2\bar{\beta}}{2\bar{\beta}+2}}\right) \text{ with } 1/\bar{\beta} = 1/(2\beta_1) + 1/(2\beta_2).$$

Notice that if $\beta_1 = \beta_2 = \beta$, thus $\overline{\beta} = \beta$ and $N^{-2\overline{\beta}/(2\overline{\beta}+2)} = N^{-2\beta/(2\beta+2)}$. The classical rate of convergence for a kernel density estimator, with one bandwidth, is $N^{-2\beta/(2\beta+1)}$. We observe here a 2 on the denominator which comes from the number of bandwidths.

In practice the regularity parameters β and R are unknown. Thus we propose a data-driven way of choosing h based on the (Goldenshluger and Lepski, 2011) method.

4.3. **Data-driven bandwidth and adaptation.** The idea is to replace the unknown term $\|f_h - f\|^2$ by an estimator. For this we introduce the iterated auxiliary kernel estimators

$$\widehat{f}_{h,h'}(x) := K_{h'} \star \widehat{f}_h(x) = \frac{1}{N}\sum_{j=1}^{N} K_{h'} \star K_h(x - \widehat{A}_j(T)).$$

Let $\mathcal{H}_{N,T}$ be a finite set of bandwidths $h = (h_1, h_2)$. Following Goldenshluger and Lepski's approach, we define the estimator of the bias term by

$$B(h) = \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\widehat{f}_{h,h'} - \widehat{f}_{h'}\|^2 - v(h')\right)_+ \tag{21}$$

where

$$v(h) := \kappa_1 \frac{\|K\|^2\|K\|_1^2}{Nh_1h_2} + \kappa_2\|K\|_1^2\left(\frac{1}{h_1^3h_2} + \frac{1}{h_1h_2^3}\right)\left(\left\|\frac{\partial K}{\partial u}\right\|^2 + \left\|\frac{\partial K}{\partial v}\right\|^2\right)\frac{C}{T}, \tag{22}$$

is a term which has the same order as the variance terms in (17), C is the constant from Proposition 1 and κ_1, κ_2 are numerical constants. The bandwidth is finally selected as follows

$$\widehat{h} = \underset{h \in \mathcal{H}_{N,T}}{\operatorname{argmin}}\, (B(h) + v(h)). \tag{23}$$

Then we obtain the following Theorem.

Theorem 6. *We assume that the elements of $\mathcal{H}_{N,T}$ satisfy $0 < h_i < 1, i = 1, 2,\ 1/(Nh_1h_2) \leq 1, 1/(h_1^3h_2) \leq T,\ 1/(h_1h_2^3) \leq T$, and that*

$$\forall c > 0,\ \exists \Sigma(c) < \infty, \qquad \sum_{h \in \mathcal{H}_{N,T}} (h_1h_2)^{-1/2} e^{-c/\sqrt{h_1h_2}} \leq \Sigma(c).$$

Under (A1)-(A5), under the assumptions of Proposition 1 and if $\|K\|_{4/3} < +\infty$, there exist numerical constants κ_1, κ_2 such that

$$\mathbb{E}[\|\widehat{f}_{\widehat{h}} - f\|^2] \leq C_1 \inf_{h \in \mathcal{H}_{N,T}} \{\|f - f_h\|^2 + v(h)\} + \frac{C_2}{N} \tag{24}$$

with C_1 a positive constant depending on $\|K\|_1$, C_2 is a positive constant depending on $\|f\|$, $\left\|\frac{\partial K}{\partial u}\right\|^2 + \left\|\frac{\partial K}{\partial v}\right\|^2$, $\|K\|_1$, $\|K\|_{4/3}$ and C (the constant from Proposition 1).

This is a non-asymptotic result proving that the bias variance compromise is automatically realized by the final estimator $\widehat{f}_{\widehat{h}}$ which is therefore adaptive. Theorem 6 is proved in Dion (2014) for a single random effect. As the proof is analogous we do not repeat it here. The key is to prove that $\mathbb{E}[B(h)] \leq \|f - f_h\|^2 + c/N$. The theoretical study gives $\kappa_1 \geq \max(40/\|K\|_1^2, 40)$ and $\kappa_2 \geq \max(10/\|K\|_1^2, 10)$. Nevertheless, the constants κ_1, κ_2 obtained in the proof are not optimal and thus it is standard to calibrate them from a simulation study.

The conditions of Theorem 6 are fulfilled for example with the set $\mathcal{H}_{N,T} = \{(1/{k_1}^2, 1/{k_2}^4), k_1 = 1\ldots, N^{1/4},\ k_2 = 1\ldots, N^{1/8},\ N^2 \leq T\}$.

In practice, note that in $v(h)$, the constant C is unknown, and must be replaced by a rough estimator. From the proofs of Propositions 1 and 7 it appears that $C = \mathbb{E}[C(\phi_j)]$ for an explicit function $C(\phi_j)$, thus C can be estimated by $(1/N)\sum_{j=1}^N C(\widehat{A}_j(T))$.

5. Discrete data.

In practice, only discrete observations are available. The estimation method must take into account this fact. Let us assume that we observe synchronously $(X_j(t)),\ j = 1, \ldots, N$ at times $t_k = k\Delta,\ k = 1, \ldots, n$ and set $T = n\Delta$. As it is classically done, we replace the stochastic and ordinary integrals by their discretized versions. The estimator of ϕ_j is thus defined by

$$\widehat{A}_{j,n} := A_{j,n}\mathbb{1}_{B_{j,n}}, \quad \text{with} \quad A_{j,n} = V_{j,n}^{-1}U_{j,n}, \tag{25}$$

and

$$U_{j,n} := \sum_{k=0}^{n-1} \frac{b}{\sigma^2}(X_j(k\Delta))(X_j((k+1)\Delta) - X_j(k\Delta)), \quad V_{j,n} := \sum_{k=0}^{n-1} \Delta \frac{b\, b^t}{\sigma^2}(X_j(k\Delta)),$$

$$B_{j,n} := \{V_{j,n} \geq \kappa\sqrt{n\Delta} I_2\}.$$

To extend the present work to discrete data, it is enough to find conditions ensuring that $\mathbb{E}[\|\widehat{A}_{j,n} - \phi_j\|_2^2] \leq \dfrac{C}{n\Delta}$. The proof follows the steps of Proposition 1, with the discretized integrals instead of the continuous ones. We must study

$$\mathbb{E}\left[\left\|\frac{1}{n\Delta}(U_{j,n} - U_j(T))\right\|_2^{2p}\right], \quad \mathbb{E}\left[\left\|\frac{1}{n\Delta}(V_{j,n} - V_j(T))\right\|_F^{2p}\right]$$

for $p = 1, 2$ and extend Proposition 7 (see Section 7) to the discretized version of $(1/T)\int_0^T g(X_j(s))ds$. This is relatively standard as we assume that the processes $(\phi_j, X_j(t))_{t\geq 0}$ are in stationary regime. The discretizations induce a new bias term which of order $O(\Delta)$. To make this bias negligible, a classical condition on the sampling interval is : $\Delta \lesssim 1/(n\Delta)$, *i.e.* $n\Delta^2 \lesssim 1$ (see Comte *et al.* (2007); Kessler *et al.* (2012) for example). Note that other approximation of stochastic or ordinary integrals based on discrete data are available and can improve the constraint $n\Delta^2 \lesssim 1$ (see *e.g.* Kloeden and Platen (1992) Iacus (2008)).

Moreover, due to the presence of random effects, additional moment conditions on the distribution of the initial condition (ϕ_j, γ_j) are required. To avoid technical developments, we do not give more details.

6. Simulation study

In this section, we illustrate on simulated data the performances of our nonparametric estimation procedure on the two examples of Sections 2.2 and 3.3. For simplicity we denote $\widehat{f} := \widehat{f}_{\widehat{h}}$ and

$$\widehat{f}_1(x_1) = \int \widehat{f}(x_1, x_2)dx_2,\ \widehat{f}_2(x_2) = \int \widehat{f}(x_1, x_2)dx_1. \tag{26}$$

For the two examples, we simulate N independent discretized sample paths $(X_j(k\Delta), k = 1, \ldots, n, j = 1, \ldots, N)$. This is done by simulating independently for each j first the random variable ϕ_j, then, given $\phi_j = \varphi$, the initial variable γ_j which is simulated according to the invariant distribution π_φ corresponding to each model. Lastly, the discretized sample path is simulated. For the OU model, we use exact simulations of the discretized path. For the CIR model, we use a discretization scheme given in Alfonsi (2005).

For simplicity we simulate independent components for $\phi_j = (\phi_{j,1}, \phi_{j,2})$, thus the density f has the form $f(x_1, x_2) = f_1(x_1) \times f_2(x_2)$.

The computation of the estimators $\widehat{A}_j(T)$ given by (13) requires to choose a value for the cut-off parameter κ. To choose κ, different functions f have been investigated with different values of Δ and

a large number of repetitions. We compared the MISEs of estimators as functions of the constant κ and selected the value $\kappa = 0.125$ as satisfactory. For comparison, we also computed the estimators for $\kappa = 0$ (no cut-off). For $\kappa = 0$ the estimator is denoted $\widetilde{f}$ and its marginals $\widetilde{f}_1$ and $\widetilde{f}_2$.
There remains the choice of the adaptive bandwidth. In Dion (2014) the kernel estimator with data-driven bandwidth selected by the Goldenshluger and Lepski criterion is implemented for one random effect. It is also implemented using the R-function *density* choosing the bandwidth by cross-validation or with the rule of thumb. The Goldenshluger and Lepski method performs better for mixture densities. Here we have chosen to use the R-function *kde2d* for the choice of the bandwidths with a standard bivariate Gaussian kernel.

Example. *[1] The simulation scheme of each sample path given φ is as follows: $X^{\varphi}(0) = \gamma \sim \mathcal{N}\left(\frac{\varphi_1}{\varphi_2}, \frac{\sigma^2}{2\varphi_2}\right)$ and $G_1, \dots, G_n$* i.i.d. *$\mathcal{N}(0,1)$ independent of $X^{\varphi}(0)$,*

$$X^{\varphi}((k+1)\Delta) = X^{\varphi}(k\Delta)e^{-\Delta\varphi_2} + \frac{\varphi_1}{\varphi_2}(1 - e^{-\Delta\varphi_2}) + \sqrt{\frac{\sigma^2}{2\varphi_2}(1 - \exp(-2\Delta\varphi_2))}G_{k+1}. \tag{27}$$

The following distributions are chosen for the components of the random effects. We drop the index j:

- $\phi_1 \sim \mathcal{N}(1, 0.5)$, $\phi_2 \sim \Gamma(10.1, 0.25)$,
- $\phi_1 \sim \mathcal{N}(1, 0.5)$, $\phi_2 \sim 1/\Gamma(3, 0.25)$
- $\phi_1 \sim \Gamma(1.8, 0.8)$, $\phi_2 \sim \Gamma(2, 1)$
- $\phi_1 \sim \Gamma(1.8, 0.8)$, $\phi_2 \sim 1/\Gamma(3, 0.25)$

Note that, for the third case, the distribution of ϕ does not satisfy all assumptions as $\mathbb{E}[\phi_2^{-10}]$ does not exist. However the simulations show that the estimation procedure works even when theoretical assumptions do not hold.

Example. *[2] The simulation scheme of each sample path given φ is the one given in Alfonsi (2005) which is based on the following approximation for small δ:*

$$X^{\varphi}_{(k+1)\delta} \cong X^{\varphi}_{k\delta} + (\varphi_1 - \sigma^2/2 - \varphi_2 X^{\varphi}_{(k+1)\delta})\delta + \sigma\sqrt{X^{\varphi}_{(k+1)\delta}}(W_{(k+1)\delta} - W_{k\delta}).$$

Solving for $X^{\varphi}((k+1)\delta)$ yields the explicit relation:

$$X^{\varphi}((k+1)\delta) = \left(\frac{\sqrt{\delta}\sigma G_{k+1} + \sqrt{\sigma^2\delta G_{k+1}^2 + 4(\delta(\varphi_1 - \sigma^2/2) + X^{\varphi}(k\delta))(1 + \varphi_2\delta)}}{2 + 2\varphi_2\delta}\right)^2 \tag{28}$$

where $G_{k+1} = (W_{(k+1)\delta} - W_{k\delta})/\sqrt{\delta}$ and $X^{\varphi}(0) = \gamma \sim \Gamma(2\frac{\varphi_1}{\sigma^2}, \frac{\sigma^2}{2\varphi_2})$. The values $\delta = T/20000$, $\Delta = 10\times\delta$ and $n = T/\Delta = 2000$ are chosen.
The two random effects must be positive random variables and the first component must satisfy $\phi_1 > 2\sigma^2$. We have chosen to simulate:

- $\phi_1 \sim 2\sigma^2 + \Gamma(5, 0.5)$, $\phi_2 \sim 1 + \Gamma(1.8, 0.8)$
- $\phi_1 \sim 2\sigma^2 + \Gamma(5, 0.5)$, $\phi_2 \sim 1/\Gamma(8.1, 0.05)$.

The chosen distribution for ϕ_1 satisfies the assumption $\mathbb{E}[\psi^2(2\phi_1/\sigma^2 - 1)] < \infty$. Indeed, when $\phi_1 \sim 2\sigma^2 + \Gamma(k, \theta)$, $\mathbb{E}[\psi^2(2\phi_1/\sigma^2 - 1)] = \int_{2\sigma^2}^{+\infty} \psi^2(2x/\sigma^2 - 1)e^{-(x-2\sigma^2)/\theta}(x - 2\sigma^2)^{k-1}dx$. One can use the asymptotic equivalent for the di-gamma function $\psi(x) \underset{+\infty}{\sim} \log(x)$ to see that the integral is convergent.
Here, the first case does not satisfy $\mathbb{E}[\phi_2^{-4}] < \infty$. But the estimation results are satisfactory.

Figures 1 and 4 illustrate the influence of the cut-off κ on the estimators for the two examples. On top, the estimators $A_j(T)$ ($\kappa = 0$) are plotted as function of the simulated variables ϕ_j. Top left is for the first component, top right for the second. The bottom figures represent the estimators $\widehat{A}_j(T)$ ($\kappa = 0.125$) as function of the simulated ϕ_j. The random variables ϕ_j are well estimated and

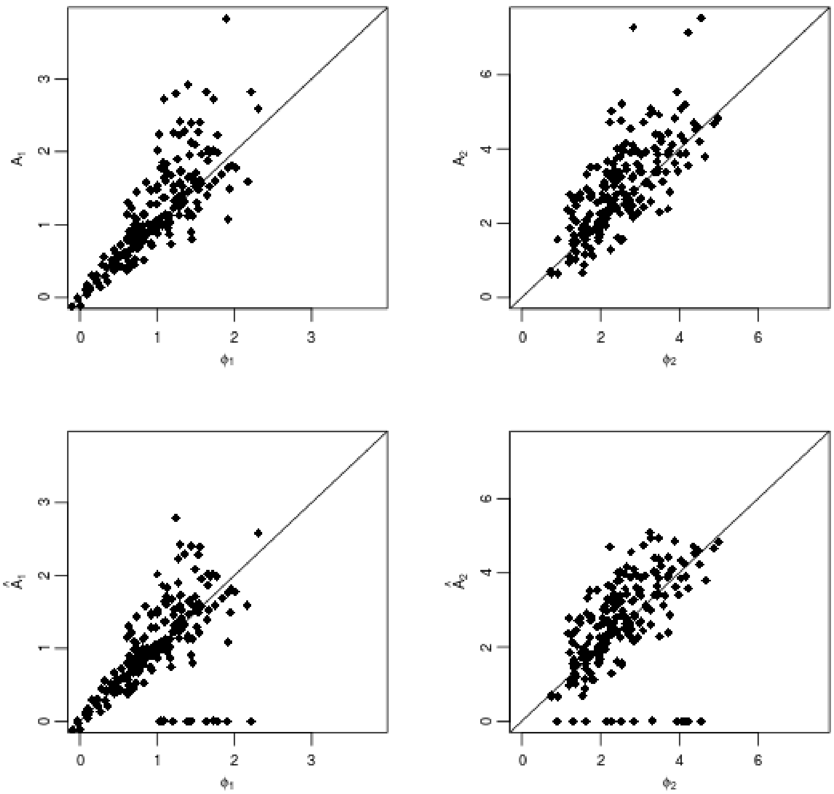

FIGURE 1. O-U example. Top left: $A_{j,1}(T)$'s as a function of $\phi_{j,1}$, top right: $A_{j,2}(T)$'s as a function of $\phi_{j,2}$. Bottom left: $\widehat{A}_{j,1}(T)$'s as a function of $\phi_{j,1}$, bottom right: $\widehat{A}_{j,2}(T)$'s as a function of $\phi_{j,2}$. With $\phi_1 \sim \mathcal{N}(1, 0.5)$, $\phi_2 \sim \Gamma(10.1, 0.25)$, $T = 10$, $N = 200$, $\sigma = 0.1$

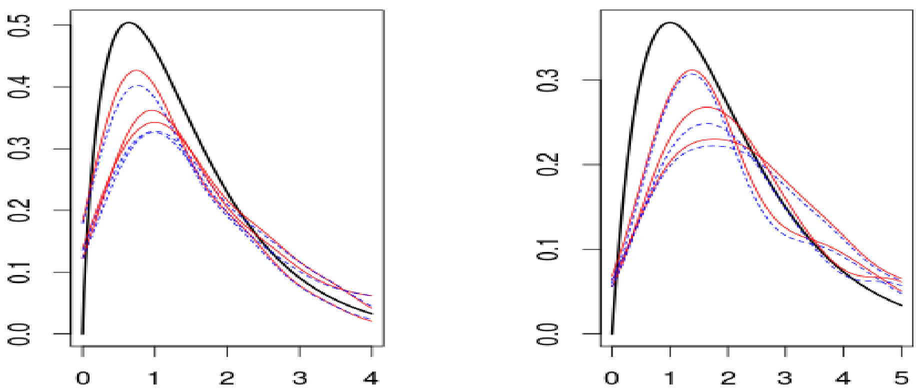

FIGURE 2. O-U example. On the left: dark bold bold line f_1, 3 estimators $\widetilde{f}_1(x_1)$ in dotted grey (blue) and $\widehat{f}_1(x_1)$ in grey (red), on the right: dark bold line f_2 3 estimators $\widetilde{f}_2(x_2)$ in dotted grey (blue) and $\widehat{f}_2(x_2)$ in grey (red). Simulation with $\phi_1 \sim \Gamma(1.8, 0.8)$, $\phi_2 \sim \Gamma(2, 1)$, $T = 10$, $N = 200$, $\sigma = 0.1$.

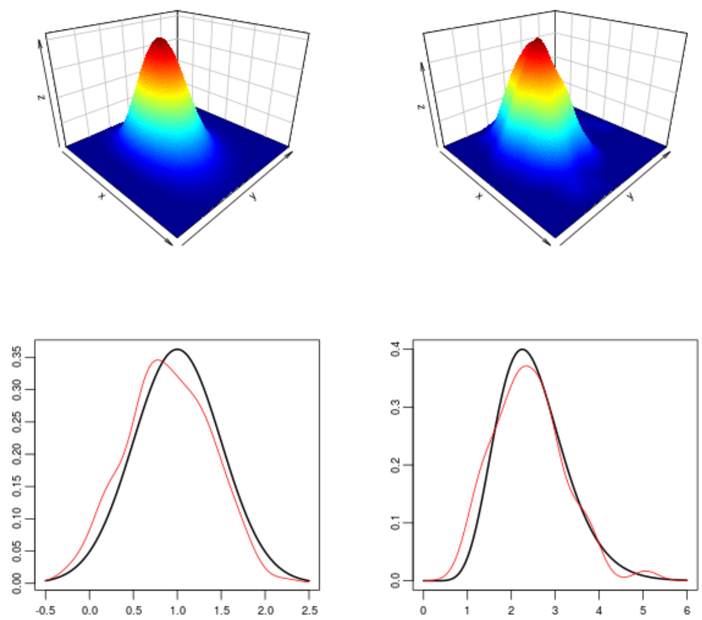

FIGURE 3. O-U example. Top left: true density f, top right: estimator $\widehat{f}$. Bottom: sections. Bottom left: dark line for $x_1 \to f(x_1, x_2)$ versus the estimator $x_1 \to \widehat{f}(x_1, x_2)$ light line (red) for x_2 fixed. Bottom right: dark line for $x_2 \to f(x_1, x_2)$ versus the estimator $x_2 \to \widehat{f}(x_1, x_2)$ light line (red), for x_1 fixed. Simulation with $\phi_1 \sim \mathcal{N}(1, 0.5)$, $\phi_2 \sim \Gamma(10.1, 0.25)$, $T = 100$, $N = 200$, $\sigma = 0.1$.

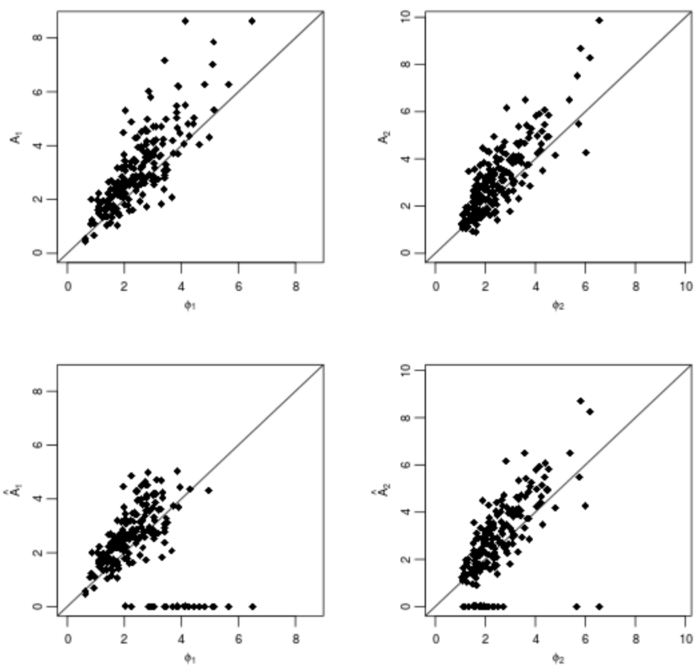

FIGURE 4. C-I-R example. Top left: $A_{j,1}(T)$'s as a function of $\phi_{j,1}$, top right $A_{j,2}(T)$'s as a function of $\phi_{j,2}$. Bottom left: $\widehat{A}_{j,1}(T)$'s as a function of $\phi_{j,1}$, top right: $\widehat{A}_{j,2}(T)$'s as a function of $\phi_{j,2}$. With $\phi_1 \sim 2\sigma^2 + \Gamma(5, 0.5)$, $\phi_2 \sim 1 + \Gamma(1.8, 0.8)$, $T = 10$, $N = 200$, $\sigma = 0.1$.

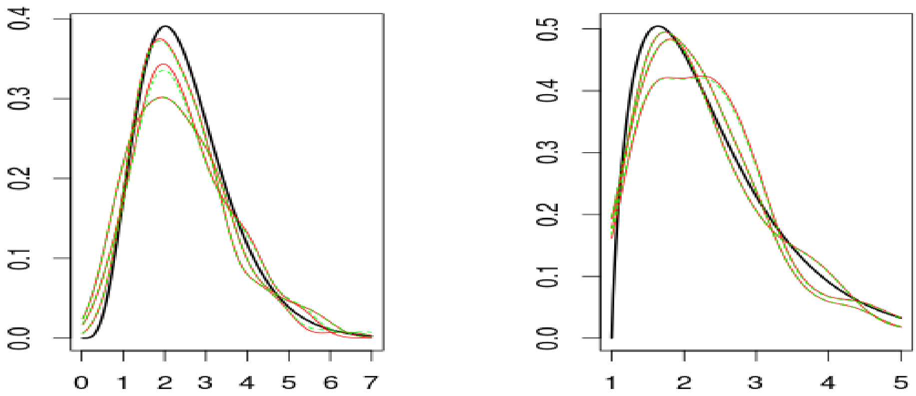

FIGURE 5. C-I-R example. On the left: dark bold line f_1, 3 estimators $\widetilde{f}_1(x_1)$ in dotted grey (green) and $\widehat{f}_1(x_1)$ in grey (red), on the right: dark bold line f_2, 3 estimators $\widetilde{f}_2(x_2)$ in dotted grey (green) and $\widehat{f}_2(x_2)$ in grey (red). Simulation with $\phi_1 \sim 2\sigma^2 + \Gamma(5, 0.5)$, $\phi_2 \sim \Gamma(1.8, 0.8)$, $T = 100$, $N = 200$, $\sigma = 0.1$.

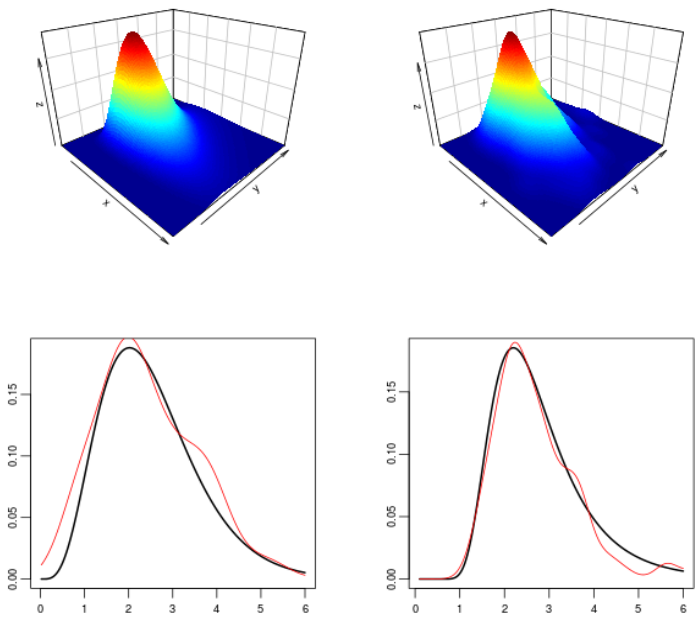

FIGURE 6. C-I-R example. Top left: true density f, top right: estimator $\widehat{f}$. Bottom: sections. Bottom left: dark line for $x_1 \to f(x_1, x_2)$ versus the estimator $x_1 \to \widehat{f}(x_1, x_2)$ light line (red) for x_2 fixed. Bottom right: dark line for $x_2 \to f(x_1, x_2)$ versus the estimator $x_2 \to \widehat{f}(x_1, x_2)$ light line (red), for x_1 fixed. Simulation with $\phi_1 \sim 2\sigma^2 + \Gamma(5, 0.5)$, $\phi_2 \sim 1/\Gamma(8.1, 0.05)$, $T = 100$, $N = 200$, $\sigma = 0.1$.

TABLE 1. O-U example. MISE computed from 100 simulations

Case	σ	T	N	$\widetilde{f}$	$\widehat{f}$	$\widehat{f_1}$	$\widehat{f_2}$
$\phi_1 \sim \mathcal{N}$, $\phi_2 \sim \Gamma$	0.1	10	20	0.064	0.064	0.067	0.073
			200	0.039	0.038	0.038	0.047
		100	20	0.046	0.046	0.047	0.030
			200	0.010	0.010	0.009	0.006
	1	100	20	0.046	0.046	0.044	0.033
			200	0.012	0.012	0.011	0.007
$\phi_1 \sim \mathcal{N}$, $\phi_2 \sim 1/\Gamma$	0.1	10	20	0.077	0.075	0.113	0.081
			200	0.050	0.051	0.008	0.060
		100	20	0.068	0.068	0.083	0.040
			200	0.021	0.027	0.035	0.010
	1	100	20	0.061	0.061	0.061	0.044
			200	0.021	0.021	0.035	0.011
$\phi_1 \sim \Gamma$, $\phi_2 \sim \Gamma$	0.1	10	20	0.028	0.024	0.056	0.047
			200	0.015	0.013	0.030	0.028
		100	20	0.025	0.025	0.029	0.027
			200	0.009	0.009	0.007	0.011
	1	100	20	0.022	0.022	0.025	0.025
			200	0.008	0.008	0.007	0.010
$\phi_1 \sim \Gamma$, $\phi_2 \sim 1/\Gamma$	0.1	10	20	0.045	0.042	0.061	0.074
			200	0.031	0.030	0.030	0.076
		100	20	0.030	0.030	0.028	0.035
			200	0.008	0.008	0.007	0.009
	1	100	20	0.029	0.029	0.027	0.034
			200	0.008	0.008	0.007	0.008

the cut-off seems to have the adequate effect: it sets to zero the values which are too far from the diagonal.

On Figures 2 (OU example) and 5 (CIR example), estimators of the two marginal densities, *i.e.* the densities of the two components of the random effects, are shown. The true marginals are in bold black. Three density estimators $\widetilde{f_1}$ (left) and $\widetilde{f_1}$ (right), built using the $A_j(T)$'s ($\kappa = 0$), are in dotted (blue) lines. Three density estimators $\widehat{f_1}$ (left), $\widehat{f_2}$ (right), built using the $\widehat{A}_j(T)$'s ($\kappa = 0.125$), are in continuous (red) lines. The cut-off for estimating the ϕ_j's improves the density estimators: especially, the first marginal $\widehat{f_1}$ fits better the true density than $\widetilde{f_1}$.

Figures 3 and 6 show the true bivariate density (top left) and the estimator $\widehat{f}$ (top right). The bottom graphs correspond to sections $x_1 \to f(x_1, x_2)$ (bold left) and $x_1 \to \widehat{f}(x_1, x_2)$ (dotted left) for a fixed x_2, and $x_2 \to f(x_1, x_2)$ (bold right) and $x_2 \to \widehat{f}(x_1, x_2)$ (dotted right) for a fixed x_1. The estimators are close to the true density.

In order to evaluate with more acuracy the performances of our estimator on the two examples, we have computed their empirical MISE from 100 simulated data sets, for different values of σ and T and for the chosen distributions of (ϕ_1, ϕ_2). Results are presented in Tables 1 and 2.

On Tables 1 and 2 the influence of the cut-off is often visible when $N = 20$: the MISEs of $\widetilde{f}$ are larger than the MISEs of $\widehat{f}$. It almost disappears on the Tables when $N = 200$. For a given value of N, when T increases from 10 to 100 the MISE decreases (twice smaller sometimes). The best results are obtained when $T = 100$ and $N = 200$. When ϕ_1 is Gaussian the estimation seems more difficult than when it is Gamma. The influence of σ may be seen when T is not very large. For example for the OU-model, the quantity σ^2/T appears is the theoretical bound of the MISE and thus it must not

Table 2. C-I-R example. MISE computed from 100 simulations

Case	σ	T	N	$\widetilde{f}$	$\widehat{f}$	$\widehat{f_1}$	$\widehat{f_2}$
$\phi_1 \sim \Gamma$, $\phi_2 \sim \Gamma$	0.1	10	20	0.031	0.030	0.049	0.080
			200	0.019	0.018	0.050	0.051
		100	20	0.022	0.021	0.027	0.031
			200	0.005	0.004	0.006	0.006
	1	100	200	0.006	0.006	0.012	0.010
$\phi_1 \sim \Gamma$, $\phi_2 \sim 1/\Gamma$	0.1	10	20	0.028	0.027	0.033	0.063
			200	0.017	0.017	0.012	0.045
		100	20	0.020	0.019	0.027	0.025
			200	0.005	0.005	0.006	0.006
	1	100	200	0.007	0.007	0.013	0.008

Table 3. CIR example. MISE computed from 100 simulations, with $N = 20$ on the left and $N = 200$ on the right, $\sigma = 0.1$, Gamma distributions for the random effects.

T	δ	Δ	T/Δ	$\widetilde{f}$	$\widehat{f}$
10	0.005	0.005×2	1000	0.038	0.037
100	0.05	0.05×2	1000	0.021	0.025
100	0.005	0.005×2	10000	0.019	0.020

T	δ	Δ	T/Δ	$\widetilde{f}$	$\widehat{f}$
10	0.005	0.005×2	1000	0.029	0.025
100	0.05	0.05×2	1000	0.006	0.007
100	0.005	0.005×2	10000	0.006	0.007

be too large. When T is large ($T = 100$) there is very little difference between the MISE's values for $\sigma = 0.1$ and $\sigma = 1$.

Table 3 highlights the different roles of T and Δ. When T increases (from 10 to 100) with the same Δ, the MISEs are divided by 2. When both T and Δ (from 0.005×2 to 0.05×2) increase, we still note a slight improvement. However, T/Δ has clearly no influence and this is why in the previous tables we fixed the number of observations per trajectory at $T/\Delta = 2000$ ($\Delta = 10 \times \delta$). The right table highlights again the role of N. [1]

7. Concluding remarks

In this work, we first provide an estimator of a bivariate random effect in the drift of a stochastic differential equation based on the observation of one trajectory given by (1). The definition of the estimator uses a cut-off parameter which allows us to study and bound its L^2-risk. Then, using the estimators of the random effects, we build a kernel estimator of the common bivariate density of the random effects, from N *i.i.d.* observed trajectories and propose a data-driven selection of the bandwidth. The Orsntein-Uhlenbeck and the Cox-Ingersoll-Ross models both with two random effects can be studied with our estimation procedure. Illustrations on simulated data are done and show the good performances of our estimator.

As these two models have many applications in finance and in neuroscience, it would be interesting to investigate our nonparametric method on real data.

The method developed here for two random effects can be easily developped for d random effects. The extension to more general drift forms would require further work and other tools to accommodate the estimation procedure, especially for estimating the random effects. Models including a linear random effect in the diffusion coefficient could be investigated, for instance using the ideas developed in Delattre *et al.* (2014).

[1] Program codes are available on the web page: `https://owncloud.math-info.univ-paris5.fr/public.php?service=files&t=bc269d7843f844a74b7cff211a751d26`

The nonparametric estimation of the drift function in SDEs with no random effect has been largely investigated in the literature (see *e.g.* Kutoyants, 2004; Comte *et al.*, 2007; Schmisser, 2013). In the case of SDEs with random effects and a general drift $b(x, \varphi_1, \varphi_2)$, the nonparametric estimation of the function b is open and of interest.

8. Proofs

8.1. **An additional result and its proof.** We denote $\mathcal{L}_\varphi$ the generator of the process given by (1) when $\phi_j = \varphi$, given for all $F \in \mathcal{C}^2((l, r))$, by $\mathcal{L}_\varphi F(x) := (\sigma^2(x)/2)F"(x) + (\,b^t(x)\varphi)F'(x)$. Then when F is a matrix, the notation $\mathcal{L}_\varphi F$ only indicated that we apply the operator on each coefficient of the matrix. We also define in this sense F' as the derivative matrix coefficients by coefficients.

Proposition 7. *Consider the processes $(X_j(t))$ given by (1) under **(A1)-(A4)**. Let $g = (g_{i,k})_{1\le i,k\le 2}$ a matrix of $S_2(\mathbb{R})$, such that $\pi_\varphi g_{i,k}^2 < +\infty$ for all $\varphi \in \Phi$. Assume that*

$$F_\varphi^g = \begin{pmatrix} F_\varphi^{g_{1,1}} & F_\varphi^{g_{1,2}} \\ F_\varphi^{g_{1,2}} & F_\varphi^{g_{2,2}} \end{pmatrix}$$

satisfies $-\mathcal{L}_\varphi F_\varphi^g = g - \pi_\varphi g$ for all $\varphi \in \Phi$. Let $J(\phi_j)$ be a non negative measurable function of ϕ_j such that for all $p \ge 1$

$$\mathbb{E}\left[J(\phi_j)\left(\frac{1}{T^p}\pi_{\phi_j}\left(\|F_{\phi_j}^g\|_F^{2p}\right) + \pi_{\phi_j}\left(\|(F_{\phi_j}^g)'\|_F^{2p}\sigma^{2p}\right)\right)\right] < \infty.$$

Then for all $p \ge 1$, there exists a constant $C_p > 0$ depending on p such that

$$\begin{aligned}\mathbb{E}\left[J(\phi_j)\left\|\sqrt{T}\left(\frac{1}{T}\int_0^T g(X_j(s))ds - \pi_{\phi_j}(g)\right)\right\|_F^{2p}\right] \;\le\; & C_p\left(\frac{1}{T^p}\mathbb{E}\left[J(\phi_j)\pi_{\phi_j}\left(\|F_{\phi_j}^g\|_F^{2p}\right)\right]\right. \\ & \left.+\mathbb{E}\left[J(\phi_j)\pi_{\phi_j}\left(\|(F_{\phi_j}^g)'\|_F^{2p}\sigma^{2p}\right)\right]\right).\end{aligned}$$

We demonstrate the result without the function J but it can be added all along. The subscript j is omitted for simplicity. Denote $g_\varphi := g - \pi_\varphi g$. Ito's formula applied to F_φ^g term by term, and the equality $\mathcal{L}_\varphi F_\varphi^g = -g_\varphi$, lead to

$$\begin{aligned}F_\varphi^g(X^\varphi(T)) &= F_\varphi^g(X^\varphi(0)) + \int_0^T \mathcal{L}_\varphi F_\varphi^g(X^\varphi(s))ds + \int_0^T (F_\varphi^g)'(X^\varphi)\sigma(X^\varphi(s))dW(s) \\ &= F_\varphi^g(X^\varphi(0)) - \int_0^T g_\varphi(X^\varphi(s))ds + \int_0^T (F_\varphi^g)'(X^\varphi(s))\sigma(X^\varphi(s))dW(s).\end{aligned}$$

Thus,

$$\int_0^T g_\varphi(X^\varphi(s))ds = \int_0^T (F_\varphi^g)'(X^\varphi(s))\sigma(X^\varphi(s))dW(s) + \left(F_\varphi^g(X^\varphi(0)) - F_\varphi^g(X^\varphi(T))\right).$$

Then with Hölder's inequality it yields

$$\mathbb{E}\left[\left\|\int_0^T g_\varphi(X^\varphi(s))ds\right\|_F^{2p}\right] \le 2^{2p-1}\left(\mathbb{T}_a + \mathbb{T}_b\right)$$

with

$$\mathbb{T}_a := \mathbb{E}\left[\left\|\int_0^T (F_\varphi^g)'(X^\varphi(s))\sigma(X^\varphi(s))dW(s)\right\|_F^{2p}\right], \quad \mathbb{T}_b := \mathbb{E}\left[\left\|F_\varphi^g(X^\varphi(0)) - F_\varphi^g(X^\varphi(T))\right\|_F^{2p}\right]. \tag{29}$$

We study first the term $\mathbb{T}_a$ given by (29). Again, Hölder's inequality gives

$$\begin{aligned}\mathbb{T}_a &= \mathbb{E}\left[\left(\sum_{1\le i,k\le 2}\left(\int_0^T (F_\varphi^{g_{i,k}})'(X^\varphi(s))\sigma(X^\varphi(s))dW(s)\right)^2\right)^p\right]\\ &\le 4^{p-1}\mathbb{E}\left[\sum_{1\le i,k\le 2}\left(\int_0^T (F_\varphi^{g_{i,k}})'(X^\varphi(s))\sigma(X^\varphi(s))dW(s)\right)^{2p}\right].\end{aligned}$$

The first term of the sum can by studied with the Burkholder-Davis-Gundy (B.D.G.) inequality (see *e.g.* Revuz and Yor, 1999; Le Gall, 2010). Thus there exists a constant $c_p > 0$ such that

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^T (F_\varphi^{g_{1,1}})'(X^\varphi(s))\sigma(X^\varphi(s))dW(s)\right)^{2p}\right] &\le c_p\mathbb{E}\left[\left(\int_0^T ((F_\varphi^{g_{1,1}})'(X^\varphi(s))\sigma(X^\varphi(s)))^2ds\right)^p\right]\\ &\le c_pT^{p-1}\mathbb{E}\left[\int_0^T ((F_\varphi^{g_{1,1}})'(X^\varphi(s))\sigma(X^\varphi(s)))^{2p}ds\right]\\ &= c_pT^{p-1}T\pi_\varphi(((F_\varphi^{g_{1,1}})'\sigma)^{2p}).\end{aligned}$$

Finally there is a constant $C_p := 4^{p-1}c_p$ verifying

$$\mathbb{T}_a \le C_pT^p\ \pi_\varphi\left(\left(\sum_{1\le i,k\le 2}\left((F_\varphi^{g_{i,k}})'\right)^{2p}\right)\sigma^{2p}\right) \le C_pT^p\ \pi_\varphi\left(\|(F_\varphi^g)'\|_F^{2p}\sigma^{2p}\right).$$

Furthermore, we study term $\mathbb{T}_b$ (29) and Cauchy-Schwarz's inequality leads to

$$\begin{aligned}\mathbb{T}_b &\le \mathbb{E}\left[\left((\|F_\varphi^g(X^\varphi(0))\|_F^{2p} + \|F_\varphi^g(X^\varphi(T))\|_F^{2p})^{1/(2p)}2^{1-1/(2p)}\right)^{2p}\right]\\ &\le 2^{2p-1}\mathbb{E}\left[\|F_\varphi^g(X^\varphi(0))\|_F^{2p} + \|F_\varphi^g(X^\varphi(T))\|_F^{2p}\right] = 2^{2p}\pi_\varphi(\|F_\varphi^g\|_F^{2p}).\end{aligned}$$

Finally, with $C'_p := \max(2^{2p}, C_p)$ it yields

$$\mathbb{E}\left[\frac{1}{T^p}\left\|\int_0^T g_\varphi(X^\varphi(s))ds\right\|_F^{2p}\right] \le C'_p\left(\pi_\varphi\left(\|(F_\varphi^g)'\|_F^{2p}\sigma^{2p}\right) + \frac{1}{T^p}\pi_\varphi(\|F_\varphi^g\|_F^{2p})\right).\ \square$$

8.2. **Proof of Proposition 1.** The subscript j is omitted for simplicity.

Proof. First note that $\widehat{A} - \phi = (A-\phi)\mathbb{1}_{B(T)} - \phi\mathbb{1}_{B(T)^c}$. Thus:

$$T\|\widehat{A}-\phi\|_2^2 = T\|(A-\phi)\mathbb{1}_{B(T)} - \phi\mathbb{1}_{B(T)^c}\|_2^2 \le 2\left(\mathbb{T}_1 + \mathbb{T}_2\right)$$

with

$$\mathbb{T}_1 := T\|(A-\phi)\|_2^2\mathbb{1}_{B(T)},\quad \mathbb{T}_2 := T\|\phi\|_2^2\mathbb{1}_{B(T)^c}. \tag{30}$$

• Study of $\mathbb{T}_1$. We have

$$\begin{aligned}\mathbb{T}_1 &= \mathbb{1}_{B(T)}T\|V(T)^{-1}M(T)\|_2^2 = \mathbb{1}_{B(T)}\left\|TV(T)^{-1}\frac{M(T)}{\sqrt{T}}\right\|_2^2\\ &= \mathbb{1}_{B(T)}\left\|(TV(T)^{-1} - L^{-1} + L^{-1})\frac{M(T)}{\sqrt{T}}\right\|_2^2 \le 2\mathbb{T}'_1 + 2\mathbb{T}_1''\end{aligned}$$

with

$$\mathbb{T}_1' := \mathbb{1}_{B(T)} \left\| (TV(T)^{-1} - L^{-1}) \frac{M(T)}{\sqrt{T}} \right\|_2^2, \quad \mathbb{T}_1\text{”} := \mathbb{1}_{B(T)} \left\| L^{-1} \frac{M(T)}{\sqrt{T}} \right\|_2^2 \tag{31}$$

Let us study this two terms separately. Note that

$$\mathbb{E}\left(\frac{V(T)}{T} \middle| \phi \right) = \frac{1}{T} \int_0^T \mathbb{E}\left(\frac{b\,b^t}{\sigma^2}(X(s)) \middle| \phi \right) ds,$$

$$\mathbb{E}\left(\frac{b\,b^t}{\sigma^2}(X(s)) \middle| \phi = \varphi \right) = \mathbb{E}\left(\frac{b\,b^t}{\sigma^2}(X^{\varphi}(s)) \right) = \mathbb{E}\left(\frac{b\,b^t}{\sigma^2}(X^{\varphi}(0)) \right) = \pi_{\varphi}\left(\frac{b\,b^t}{\sigma^2} \right).$$

Thus

$$L := \pi_{\phi}\left(\frac{b\,b^t}{\sigma^2} \right) = \mathbb{E}\left(\frac{V(T)}{T} \middle| \phi \right).$$

• Study of $\mathbb{T}_1$”. As L is $\mathcal{F}_0$ measurable,

$$\begin{aligned} \mathbb{E}[\mathbb{T}_1\text{”}] &\leq \mathbb{E}\left[\|L^{-1}\|_F^2 \left\| \frac{M(T)}{\sqrt{T}} \right\|_2^2 \right] = \mathbb{E}\left[\|L^{-1}\|_F^2 \mathbb{E}\left[\left(\frac{M_1^2(T)}{T} + \frac{M_2^2(T)}{T} \right) \middle| \mathcal{F}_0 \right] \right] \\ &= \mathbb{E}\left[\|L^{-1}\|_F^2 \left(\frac{< M_1 >_T}{T} + \frac{< M_2 >_T}{T} \right) \right] = \mathbb{E}\left[\|L^{-1}\|_F^2 \mathrm{Tr}\left(\frac{V(T)}{T} \right) \right] \end{aligned}$$

Thus we obtain:

$$\mathbb{E}[\mathbb{T}_1\text{”}] \leq \mathbb{E}\left[\|L^{-1}\|_F^2 \mathbb{E}\left[\mathrm{Tr}\left(\frac{V(T)}{T} \right) \middle| \phi \right] \right] = \mathbb{E}\left[\|L^{-1}\|_F^2 \mathrm{Tr}(L) \right].$$

Using Property 9 (see Appendix), we must assume :

$$\mathbb{E}\left[\frac{\|L\|_F^2}{\det(L)^2} \mathrm{Tr}\,(L) \right] < \infty. \tag{32}$$

• Study of $\mathbb{T}_1'$ given by (31). This term will use the cut-off $B(T)$. We have:

$$\begin{aligned} TV(T)^{-1} - L^{-1} &= (TV(T)^{-1}L - I)L^{-1} \\ &= V(T)^{-1}(TL - V(T))L^{-1}. \end{aligned} \tag{33}$$

Then we obtain:

$$\begin{aligned} \left(TV(T)^{-1} - L^{-1}\right) \frac{M(T)}{\sqrt{T}} &= V(T)^{-1}(TL - V(T))L^{-1}\frac{M(T)}{\sqrt{T}} \\ &= \underbrace{TV(T)^{-1}\left(L - \frac{V(T)}{T} \right)}_{\text{matrix}} \underbrace{L^{-1}\frac{M(T)}{\sqrt{T}}}_{\text{vector}} \end{aligned}$$

and

$$\mathbb{E}[\mathbb{T}_1'] \leq \mathbb{E}\left[\|TV^{-1}(T)\|_F^2 \mathbb{1}_{B(T)} \left\| L - \frac{V(T)}{T} \right\|_F^2 \left\| L^{-1}\frac{M(T)}{\sqrt{T}} \right\|_2^2 \right].$$

But $\|V(T)^{-1}\|_F^2 = \det(V(T)^{-2})\|V(T)\|_F^2 = \lambda_1(T)^{-2} + \lambda_2(T)^{-2}$. And the definition of the set $B(T)$ implies:

$$\|TV(T)^{-1}\|_F^2 \leq \frac{2T}{\kappa^2}. \tag{34}$$

Thus the Hölder inequality yields

$$\mathbb{E}[\mathbb{T}_1'] \leq \mathbb{E}\left[\frac{2T}{\kappa^2} \left\| L - \frac{V(T)}{T} \right\|_F^2 \left\| L^{-1}\frac{M(T)}{\sqrt{T}} \right\|_2^2 \right] \leq \frac{2}{\kappa^2} \left(\mathbb{E}[\mathbb{T}_c]\mathbb{E}[\mathbb{T}_d] \right)^{1/2}$$

with

$$\mathbb{T}_c := \left\|\sqrt{T}\left(L - \frac{V(T)}{T}\right)\right\|_F^4, \quad \mathbb{T}_d := \left\|L^{-1}\frac{M(T)}{\sqrt{T}}\right\|_2^4. \tag{35}$$

• Study of term $\mathbb{T}_c$. We apply here Proposition 7 with $p = 2$ and $g = b\,^t b/\sigma^2$ and get

$$\mathbb{E}\left[\mathbb{T}_c\right] \le C\left(\frac{1}{T^2}\mathbb{E}\left[\|H_\phi\|_F^4\right] + \mathbb{E}\left[\|H'_\phi\|_F^4\sigma^4\right]\right) < \infty. \tag{36}$$

• Study of term $\mathbb{T}_d$. We notice that:

$$\mathbb{E}[\mathbb{T}_d] \le \mathbb{E}\left[\|L^{-1}\|_F^4\left\|\frac{M(T)}{\sqrt{T}}\right\|_2^4\right].$$

Proceeding as for $\mathbb{T}_1$" and using the B.D.G. inequality implies that for some constant $c > 0$,

$$\begin{aligned}\mathbb{E}\left[\|L^{-1}\|_F^4\left\|\frac{M(T)}{\sqrt{T}}\right\|_2^4\right] &= \mathbb{E}\left[\|L^{-1}\|_F^4\left(\frac{M_1^2(T)+M_2^2(T)}{T}\right)^2\right] \\ &\le c\,\mathbb{E}\left[\|L^{-1}\|_F^4\mathbb{E}\left[\left(\mathrm{Tr}\left(\frac{V(T)}{T}\right)\right)^2\middle|\phi\right]\right].\end{aligned}$$

And by the Hölder inequality,

$$\begin{aligned}\mathbb{E}\left[\left(\mathrm{Tr}\left(\frac{V(T)}{T}\right)\right)^2\middle|\phi=\varphi\right] &= \mathbb{E}\left[\left(\frac{1}{T}\int_0^T\left(\frac{b_1^2}{\sigma^2}(X^\varphi(s))+\frac{b_2^2}{\sigma^2}(X^\varphi(s))\right)ds\right)^2\right] \\ &\le \mathbb{E}\left[\frac{1}{T}\int_0^T\left(\frac{b_1^2}{\sigma^2}(X^\varphi(s))+\frac{b_2^2}{\sigma^2}(X^\varphi(s))\right)^2 ds\right] = \mathrm{Tr}\left(\pi_\varphi\left(\left(\frac{b\,b^t}{\sigma^2}\right)^2\right)\right).\end{aligned}$$

Thus,

$$\mathbb{E}\left[\mathbb{T}_d\right] \le \mathbb{E}\left[\|L^{-1}\|_F^4\mathrm{Tr}\left(\pi_\phi\left(\left(\frac{b\,b^t}{\sigma^2}\right)^2\right)\right)\right].$$

Finally we must assume that

$$\mathbb{E}\left[\frac{\|L\|_F^4}{\det(L)^4}\mathrm{Tr}\left(\pi_\phi\left(\left(\frac{b\,b^t}{\sigma^2}\right)^2\right)\right)\right] < \infty. \tag{37}$$

• Study of term $\mathbb{T}_2$. Term $\mathbb{T}_2$ is given by (30). Note that

$$\mathbb{E}[\mathbb{T}_2] = T\mathbb{E}[\mathbb{E}[\|\phi\|_2^2\,\mathbb{1}_{\{B(T)^c\}}|\,\phi]] = T\mathbb{E}[\|\phi\|_2^2\mathbb{E}[\mathbb{1}_{\{B(T)^c\}}|\,\phi]].$$

We need to compute the quantity $\mathbb{P}(B(T)^c|\phi=\varphi)]$ for all φ. We have:

$$B(T)^c = \left\{\frac{V(T)}{T} \le \frac{\kappa}{\sqrt{T}}I_2\right\} = \left\{L - \frac{V(T)}{T} \ge L - \frac{\kappa}{\sqrt{T}}I_2\right\}.$$

If we denote $l := \pi_\varphi\left(\frac{b\,b^t}{\sigma^2}\right)$,

$$l - \frac{\kappa}{\sqrt{T}}I_2 \ge \frac{l}{2} \Leftrightarrow l - \frac{l}{2} \ge \frac{\kappa}{\sqrt{T}}I_2 \Leftrightarrow l \ge \frac{2\kappa}{\sqrt{T}}I_2.$$

Furthermore,

$$\begin{aligned}\mathbb{P}\left(\frac{V(T)}{T} \le \frac{\kappa}{\sqrt{T}} I_2 \middle| \phi = \varphi\right) &\le \mathbb{1}_{l \ge \frac{2\kappa}{\sqrt{T}} I_2} \mathbb{P}\left(\frac{V(T)}{T} \le \frac{\kappa}{\sqrt{T}} I_2 \middle| \phi = \varphi\right) + \mathbb{1}_{l \le \frac{2\kappa}{\sqrt{T}} I_2} \\ &\le \mathbb{T}_\alpha + \mathbb{T}_\beta\end{aligned}$$

with

$$\mathbb{T}_\alpha := \mathbb{P}\left(l - \frac{V(T)}{T} \ge \frac{l}{2} \middle| \phi = \varphi\right), \quad \mathbb{T}_\beta := \mathbb{1}_{l \le \frac{2\kappa}{\sqrt{T}} I_2}. \tag{38}$$

We study the two terms separately. First:

$$\mathbb{T}_\beta = \mathbb{1}_{l \le \frac{2\kappa}{\sqrt{T}} I_2} = \mathbb{1}_{l^{-1} \ge \frac{\sqrt{T}}{2\kappa} I_2} = \mathbb{1}_{\min\{\frac{1}{\lambda_{\varphi,1}}, \frac{1}{\lambda_{\varphi,2}}\} \ge \frac{\sqrt{T}}{2\kappa}} \le \left(\frac{4\kappa^2}{T}\right) \min\left\{\frac{1}{\lambda_{\varphi,1}}, \frac{1}{\lambda_{\varphi,2}}\right\}$$

Secondly, for the term $\mathbb{T}_\alpha$ given by formula (38), we use Property 11 (Appendix)

$$l - \frac{V(T)}{T} \ge \frac{l}{2} \Rightarrow l_{i,i} - \frac{V_{i,i}(T)}{T} \ge \frac{l_{i,i}}{2}, \ i = 1, 2$$

then

$$\mathbb{T}_\alpha \le \mathbb{P}\left(l_{1,1} - \frac{V_{1,1}(T)}{T} \ge \frac{l_{1,1}}{2} \middle| \phi = \varphi\right) + \mathbb{P}\left(l_{2,2} - \frac{V_{2,2}(T)}{T} \ge \frac{l_{2,2}}{2} \middle| \phi = \varphi\right).$$

For the two terms of the right hand side we are able now to use Markov inequality:

$$\mathbb{T}_\alpha \le \left(\frac{2}{l_{1,1}}\right)^2 \mathbb{E}\left[\left(l_{1,1} - \frac{V_{1,1}(T)}{T}\right)^2 \middle| \phi = \varphi\right] + \left(\frac{2}{l_{2,2}}\right)^2 \mathbb{E}\left[\left(l_{2,2} - \frac{V_{2,2}(T)}{T}\right)^2 \middle| \phi = \varphi\right].$$

Finally we obtain

$$\begin{aligned}\mathbb{E}[T_2] &\le T\mathbb{E}\left[\|\phi\|_2^2 \left(\left(\frac{2}{L_{1,1}}\right)^2 \left(L_{1,1} - \frac{V_{1,1}(T)}{T}\right)^2 + \left(\frac{2}{L_{2,2}}\right)^2 \left(L_{2,2} - \frac{V_{2,2}(T)}{T}\right)^2\right.\right. \qquad (39) \\ &\qquad \left.\left. + \left(\frac{4\kappa^2}{T}\right) \min\left\{\frac{1}{\lambda_1}, \frac{1}{\lambda_2}\right\}\right)\right] \\ &\le \mathbb{E}\left[4\|\phi\|_2^2 \left(\frac{1}{L_{1,1}^2} + \frac{1}{L_{2,2}^2}\right) \left\|\sqrt{T}\left(L - \frac{V(T)}{T}\right)\right\|^2\right] + \mathbb{E}\left[4\kappa^2 \|\phi\|_2^2 \min\left\{\frac{1}{\lambda_1}, \frac{1}{\lambda_2}\right\}\right] \qquad (40)\end{aligned}$$

We are able to conclude this proof using the Proposition 7 with $J(\phi) = \|\phi\|_2^2 \left(\frac{1}{L_{1,1}^2} + \frac{1}{L_{2,2}^2}\right)$. Indeed, as $T\mathbb{E}[\|\widehat{A} - \phi\|^2] \le 4\mathbb{E}[\mathbb{T}_1']4\mathbb{E}[\mathbb{T}_1"] + 2\mathbb{E}[\mathbb{T}_2]$, gathering the bounds (32), (36), (37),and (39) we obtain that there is a constant $C > 0$ such that $\mathbb{E}[\|\widehat{A}_j(T) - \phi_j\|_2^2] \le C/T$.

□

8.3. **Proof of proposition 2.**

Proof. With the Cauchy-Schwarz inequality, we obtain the following decomposition

$$\begin{aligned}\mathbb{E}[\|\widehat{f}_h - f\|^2] &= \|f - \mathbb{E}[\widehat{f}_h]\|^2 + \mathbb{E}[\|\mathbb{E}[\widehat{f}_h] - \widehat{f}_h\|^2] \\ &\le 2\|f - f_h\|^2 + 2\|f_h - \mathbb{E}[\widehat{f}_h]\|^2 + \mathbb{E}[\|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|^2].\end{aligned}$$

Let us study the second term of the bound. For all $x = (x_1, x_2) \in \mathbb{R}^2$,

$$\mathbb{E}[\widehat{f}_h(x)] = \frac{1}{N}\sum_{j=1}^{N} \mathbb{E}[K_h(x - \widehat{A}_j(T))] = \mathbb{E}[K_h(x - \widehat{A}_1(T))],$$

thus $f_h \neq \mathbb{E}[\widehat{f}_h]$. The subscript j (or 1) is omitted in the following for simplicity. Note that

$$f_h(x) = f_h(x_1, x_2) \quad = \quad f \star K_h(x_1, x_2) = \frac{1}{h_1 h_2} \int_{\mathbb{R}^2} f(u, v) K\left(\frac{x_1 - u}{h_1}, \frac{x_2 - v}{h_2}\right) du dv = \mathbb{E}[K_h(x - \phi)].$$

Then to study $\|f_h - \mathbb{E}[\widehat{f}_h]\|^2 = \int_{\mathbb{R}^2} \mathbb{E}[K_h(x - \widehat{A}(T)) - K_h(x - \phi)]^2 dx$ we denote $DK_h = (\frac{\partial K_h}{\partial u}, \frac{\partial K_h}{\partial v})^t$, with

$$\frac{\partial K_h}{\partial u}(u, v) = \frac{1}{h_1^2 h_2} \frac{\partial K}{\partial u}\left(\frac{u}{h_1}, \frac{v}{h_2}\right), \; \frac{\partial K_h}{\partial v}(u, v) = \frac{1}{h_1 h_2^2} \frac{\partial K}{\partial v}\left(\frac{u}{h_1}, \frac{v}{h_2}\right)$$

and applying Taylor's formula with $U(x, t) := (U_1(x, t), U_2(x, t)) := x - \phi + t(\phi - \widehat{A}(T)$, it yields

$$K_h(x - \widehat{A}(T)) - K_h(x - \phi) \quad = \quad \int_0^1 DK_h(U(x, t)).(\phi - \widehat{A}(T)) dt.$$

Thus

$$\begin{aligned}
\|f_h - \mathbb{E}[\widehat{f}_h]\|^2 \quad &\leq \quad \int \mathbb{E}\left[(K_h(x - \phi) - K_h(x - \widehat{A}(T)))^2\right] dx = \mathbb{E}\left[\int (K_h(x - \phi) - K_h(x - \widehat{A}(T)))^2 dx\right] \\
&\leq \quad \mathbb{E}\left[2 \int_{\mathbb{R}^2} \left(\int_0^1 \frac{2}{h_1^2 h_2} \frac{\partial K}{\partial u}\left(\frac{U_1(x, t)}{h_1}, \frac{U_2(x, t)}{h_2}\right) dt\right)^2 dx_1 dx_2 \; (\phi_1 - \widehat{A}_1(T))^2\right] \\
&\quad + \mathbb{E}\left[2 \int_{\mathbb{R}^2} \left(\int_0^1 \frac{2}{h_1 h_2^2} \frac{\partial K}{\partial v}\left(\frac{U_1(x, t)}{h_1}, \frac{U_2(x, t)}{h_2}\right) dt\right)^2 dx_1 dx_2 \; (\phi_2 - \widehat{A}_2(T))^2\right].
\end{aligned}$$

The first term of the previous sum is bounded according to Proposition 1:

$$\begin{aligned}
&\mathbb{E} \quad \left[\int_{\mathbb{R}^2} \left(\int_0^1 \frac{1}{h_1^2 h_2} \frac{\partial K}{\partial u}\left(\frac{U_{1,1}}{h_1}, \frac{U_{1,2}}{h_2}\right) dt\right)^2 dx_1 dx_2 \; (\phi_{1,1} - \widehat{A}_{1,1}(T))^2\right] \\
&\leq \quad \mathbb{E}\left(\frac{1}{h_1^4 h_2^2} \int_0^1 dt \int_{\mathbb{R}^2} \left(\frac{\partial K}{\partial u}(y_1, y_2)\right)^2 dy_1 dy_2 \; h_1 h_2 \; (\phi_{1,1} - \widehat{A}_{1,1}(T))^2\right) \\
&\leq \quad \frac{1}{h_1^3 h_2} \int_{\mathbb{R}^2} \left(\frac{\partial K}{\partial u}(y_1, y_2)\right)^2 dy_1 dy_2 \; \mathbb{E}[(\phi_{1,1} - \widehat{A}_{1,1}(T))^2] \leq \frac{C}{h_1^3 h_2 \; T} \left\|\frac{\partial K}{\partial u}\right\|^2
\end{aligned}$$

with C the constant from Equation (15). The same arguments works for the second term and finally it yields

$$\|f_h - \mathbb{E}[\widehat{f}_h]\|^2 \quad \leq \quad C \max\left(\frac{1}{h_1^3 h_2}, \frac{1}{h_1 h_2^3}\right) \frac{\left\|\frac{\partial K}{\partial u}\right\|^2 + \left\|\frac{\partial K}{\partial v}\right\|^2}{T}.$$

Finally, the last term is bounded by

$$\mathbb{E}[\|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|^2] \leq \frac{\|K\|^2}{N h_1 h_2}.$$

$\square$

8.4. **Proof of Proposition 4.** For all $x = (x_1, x_2) \in \mathbb{R}^2$,

$$f_h(x) - f(x) = f \star K_h(x) - f(x) = \int_{\mathbb{R}^2} K(v)[f(x - vh) - f(x)] dv.$$

Then,

$$f(x_1 - v_1 h_1, x_2 - v_2 h_2) - f(x_1, x_2) = f(x_1 - v_1 h_1, x_2 - v_2 h_2) - f(x_1 - v_1 h_1, x_2) + f(x_1 - v_1 h_1, x_2) - f(x_1, x_2).$$

As in Tsybakov (2009) we apply Taylor's formula to the two partial functions: $t \mapsto f(x_1 - v_1 h_1, t)$ at the order l_2 and $t \mapsto f(t, x_2)$ at the order l_1. Using the orders of the two kernels yields for the first term,

$$
\begin{aligned}
B_1(x) &:= \int_{\mathbb{R}^2} K_1(v_1)K_2(v_2)[f(x_1-v_1h_1,x_2-v_2h_2)-f(x_1-v_1h_1,x_2)]dv_1dv_2 \qquad (41)\\
&= -\int_{\mathbb{R}^2}\frac{(h_2v_2)^{l_2}}{(l_2-1)!}\int_0^1(1-t)^{l_2-1}\frac{\partial^{l_2}}{\partial x_2^{l_2}}f(x_1-v_1h_1,x_2-v_2h_2+th_2v_2)dt\ K_1(v_1)K_2(v_2)dv_1dv_2\\
&= -\int_{\mathbb{R}^2}\frac{(h_2v_2)^{l_2}}{(l_2-1)!}\int_0^1(1-t)^{l_2-1}\left[\frac{\partial^{l_2}}{\partial x_2^{l_2}}f(x_1-v_1h_1,x_2-v_2h_2+th_2v_2)\right.\\
&\qquad\left.-\frac{\partial^{l_2}}{\partial x_2^{l_2}}f(x_1-v_1h_1,x_2-v_2h_2)\right]dtK_1(v_1)K_2(v_2)dv_1dv_2
\end{aligned}
$$

and the analogue term $B_2(x)$. To evaluate $\|f-f_h\|^2=\int(f_h(x)-f(x))^2dx$ we remind the generalized Minkowski's inequality (see Tsybakov, 2009): for any measurable function f on $\mathbb{R}^2$, we have

$$
\int\left(\int f(u,x)du\right)^2dx\le\left[\int\left(\int f^2(u,x)dx\right)^{1/2}du\right]^2.
$$

Looking at the first term (41), and applying twice the above inequality implies for $f\in\mathcal{N}(\beta,R)$:

$$
\begin{aligned}
\int_{\mathbb{R}^2}B_1^2(x)dx &\le \left(\int_{\mathbb{R}^2}|K_1(v_1)K_2(v_2)|\frac{h_2|v_2|^{l_2}}{(l_2-1)!}\int_0^1(1-t)^{l_2-1}\left[\int_{\mathbb{R}^2}\left(\frac{\partial^{l_2}}{\partial x_2^{l_2}}f(x_1-v_1h_1,x_2-v_2h_2+t\ h_2v_2)\right.\right.\right.\\
&\qquad\left.\left.\left.-\frac{\partial^{l_2}}{\partial x_2^{l_2}}f(x_1-v_1h_1,x_2-v_2h_2)\right)^2dx_1dx_2\right]^{1/2}dt\ dv_1dv_2\right)^2\\
&\le \left[\int_{\mathbb{R}^2}|K_1(v_1)K_2(v_2)|{h_2}^{\beta_2}\frac{|v_2|^{\beta_2}}{l_2!}R\ dv_1dv_2\right]^2\le C_2^2h_2^{2\beta_2}
\end{aligned}
$$

with $C_2:=(R/l_2!)\int|K_2(v_2)v_2^{\beta_2}|dv_2$. Finally,

$$
\begin{aligned}
\int(f_h(x)-f(x))^2dx &\le 2\int_{\mathbb{R}^2}B_1^2(x)dx+2\int_{\mathbb{R}^2}B_2^2(x)dx\\
&\le 2C_1^2h_1^{2\beta_1}+2C_2^2h_2^{2\beta_2}.\ \square
\end{aligned}
$$

Appendix

8.5. Useful results of algebra.

Property 8. *For all $A,B\in M_2(\mathbb{R})$, $\|AB\|_F\le\|A\|_F\|B\|_F$ and for all $x\in\mathbb{R}^2$, $\|Ax\|_2\le\|A\|_F\|x\|_2$ where $\|\|_2$ is the euclidean norm on $\mathbb{R}^2$.*

Property 9. *If $A\in S_2(\mathbb{R})$ and $(\lambda_1,\lambda_2)\in\mathbb{R}^2$ are its eigenvalues, $\|A\|_F^2=\lambda_1^2+\lambda_2^2$. Furthermore if A is invertible, $(\lambda_1,\lambda_2)\in\mathbb{R}^{*2}$ and $\|A^{-1}\|_F^2=\left(\frac{1}{\det(A)}\right)^2\|A\|_F^2=\frac{1}{\lambda_1^2}+\frac{1}{\lambda_2^2}$.*

Property 10. *If $A\in S_2(\mathbb{R})$ with eigenvalues λ_1,λ_2, then $A\ge I_2\Leftrightarrow\min(\lambda_1,\lambda_2)\ge 1$.*

Property 11. *For all $A\in S_2^+(\mathbb{R})$, $a_{i,i}\ge 0$ for $i=1,2$. Then if $A,B\in S_2(\mathbb{R})$,*

$$
A\ge B\Rightarrow\ a_{i,i}\ge b_{i,i}\quad i=1,2.
$$

8.6. **Details on examples.** We refer to Genon-Catalot *et al.* (2000) and Genon-Catalot and Larédo (2014) for details and properties on the generator infinitesimal. Nevertheless we recall one of them. For $f \in \mathcal{C}^2((l,r))$,

$$\mathcal{L}_\varphi f(x) = \frac{1}{2}\sigma^2(x) f''(x) + ({}^t b(x)\varphi) f'(x) = \frac{1}{2m_\varphi(x)} \left(\frac{f'}{s_\varphi}\right)'(x). \tag{42}$$

Thus for $g \in L^2_{\pi_\varphi}$ when we are looking for the associated F^g_φ such that $\mathcal{L}_\varphi F^g_\varphi = -(g - \pi_\varphi(g))$ we can use the relation $(F^g_\varphi)'(x) = -2s_\varphi(x)\int_l^x (g(u) - \pi_\varphi(g)) m_\varphi(u) du$.

For the O-U model.
We want to explicit $H_\phi = F_\phi^{b\,b^t/\sigma^2}$. Let us denote: $g = b\,b^t/\sigma^2 = (g_{i,k})_{1\le i,k\le 2}$. First: $g_{1,1}(x) - \pi_\varphi g_{1,1} = 0 = -\mathcal{L}_\varphi F^{g_{1,1}}_\varphi(x)$. For example $F^{g_{1,1}}_\varphi(x) = 1$ is suitable. Then, $g_{1,2}(x) - \pi_\varphi g_{1,2} = -\frac{1}{\sigma^2}\left(x - \frac{\varphi_1}{\varphi_2}\right)$. We look for $F^{g_{1,2}}_\varphi(x) = -\frac{a}{\sigma^2}\left(x - \frac{\varphi_1}{\varphi_2}\right)$ and we obtain

$$F^{g_{1,2}}_\varphi(x) = -\frac{1}{\varphi_2\sigma^2}\left(x - \frac{\varphi_1}{\varphi_2}\right).$$

Finally

$$g_{2,2}(x) - \pi_\varphi g_{2,2} = \frac{x^2}{\sigma^2} - \frac{1}{2\varphi_2} + \frac{\varphi_1^2}{\sigma^2\varphi_2^2} = \frac{1}{\sigma^2}\left[\left(x - \frac{\varphi_1}{\varphi_2}\right)^2 + 2\frac{\varphi_1}{\varphi_2}\left(x - \frac{\varphi_1}{\varphi_2}\right)\right] - \frac{1}{2\varphi_2}.$$

We look for $F^{g_{2,2}}_\varphi$ with the same functional form and we obtain

$$F^{g_{2,2}}_\varphi(x) = \frac{2\varphi_1}{\sigma^2\varphi_2^2}\left(x - \frac{\varphi_1}{\varphi_2}\right) + \frac{1}{2\sigma^2\varphi_2}\left[\left(x - \frac{\varphi_1}{\varphi_2}\right)^2 - \frac{\sigma^2}{2\varphi_2}\right].$$

For the C-I-R model.
We want to explicit $H_\phi = F_\phi^{b\,b^t/\sigma^2}$. First: $g_{1,1}(x) - \pi_\varphi g_{1,1} = 1/(\sigma^2 x) - 2\varphi_2/((2\varphi_1 - \sigma^2)\sigma^2)$. Here we use formula (42). We have

$$(F^{g_{1,1}}_\varphi)'(x) = -2s_\varphi(x)\int_0^x \left(\frac{1}{\sigma^2 u} - \frac{2\varphi_2}{\sigma^2(2\varphi_1 - \sigma^2)}\right) m_\varphi(u) du$$

with $s_\varphi(x) = e^{2\varphi_2 x/\sigma^2} x^{-2\varphi_1/\sigma^2}$. This is equivalent to

$$(F^{g_{1,1}}_\varphi)'(x) = -\frac{2}{\sigma^4} e^{2\varphi_2 x/\sigma^2} x^{-2\varphi_1/\sigma^2}\left[\int_0^x e^{-2\varphi_2 u/\sigma^2} u^{2\varphi_1/\sigma^2 - 2} du - \int_0^x \frac{2\varphi_2}{2\varphi_1 - \sigma^2} e^{-2\varphi_2 u/\sigma^2} u^{2\varphi_1/\sigma^2 - 1} du\right]$$

and an integration by part in the first integral gives

$$(F^{g_{1,1}}_\varphi)'(x) = -\frac{2}{\sigma^4} e^{2\varphi_2 x/\sigma^2} x^{-2\varphi_1/\sigma^2}\left[-\frac{e^{-2\varphi_2 x/\sigma^2} x^{2\varphi_1/\sigma^2 - 1}}{2\varphi_1/\sigma^2 - 1}\right] = \frac{2}{\sigma^2(2\varphi_1 - \sigma^2)x}.$$

We finally set

$$F^{g_{1,1}}_\varphi(x) = \frac{2}{\sigma^2(2\phi_1 - \sigma^2)}\log(x).$$

Furthermore $g_{1,2}(x) - \pi_\varphi g_{1,2} = 0 = -\mathcal{L}_\varphi F^{g_{1,2}}_\varphi(x)$, and then for example $F^{g_{1,2}}_\varphi(x) = 1$. At last $g_{2,2}(x) - \pi_\varphi g_{2,2} = \frac{1}{\sigma^2}(x - \frac{\varphi_1}{\varphi_2})$ and we obtain

$$F^{g_{2,2}}_\varphi(x) = \frac{1}{\varphi_2\sigma^2}\left(x - \frac{\varphi_1}{\varphi_2}\right).$$

References

Alfonsi, A. (2005). On the discretization schemes for the cir (and bessel squared) processes. *Monte Carlo Methods and Applications* **11**, 355–384.

Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13**, 514–543.

Comte, F., Genon-Catalot, V. and Samson, A. (2013). Nonparametric estimation for stochastic differential equation with random effects. *Stochastic Processes and their Applications* **7**, 2522–2551.

Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques* **49**, 569–609.

Davidian, M. and Giltinan, D. (1995). *Nonlinear models for repeated measurement data.* Chapman and Hall, London.

Delattre, M., Genon-Catalot, V. and Samson, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics* **40**, 322–343.

Delattre, M., Genon-Catalot, V. and Samson, A. (2014). Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient. *hal-01056917, MAP5 2014-07* .

Delattre, M. and Lavielle, M. (2013). Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Statistics and Its Interface* **6**, 519–532.

Diggle, P., Heagerty, P., Liang, K. and Zeger, S. (2002). *Analysis of longitudinal data.* Oxford University Press.

Dion, C. (2014). Nonparametric estimation in a mixed-effect Ornstein-Uhlenbeck model. *hal-01023300, MAP5 2014-24* .

Donnet, S. and Samson, A. (2013). A review on estimation of stochastic differential equations for pharmacokinetic - pharmacodynamic models. *Advanced Drug Delivery Reviews* **65**, 929–939.

Genon-Catalot, V., Jeantheau, T. and Laredo, C. (2000). Stochastic volatility models as hidden markov models and statistical applications. *Bernoulli* **6**, 1051–1079.

Genon-Catalot, V. and Larédo, C. (2014). Estimation for stochastic differential equations with mixed effects. *hal-00807258 v2* .

Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. *The Annals of Statistics* **39**, 1608–1632.

Iacus, S. (2008). *Simulation and Inference for stochastic differential equation.* Springer Series in Statistics.

Kerkyacharian, G., Lepski, O. and Picard, D. (2007). Nonlinear estimation in anisotropic multiindex denoising. *Theory Probab. Appl.* **52**, 150–171.

Kessler, M., Lindner, A. and Sorensen, M. (2012). *Statistical methods for stochastic differential equations*, vol. 124. CRC Press.

Kloeden, P. and Platen, E. (1992). *Numerical solution of stochastic differential equations.* Springer, Berlin.

Kutoyants, Y. (2004). *Statistical Inference for Ergodic Diffusion Processes.* Springer, London.

Le Gall, J. (2010). Calcul stochastique et processus de markov. *Notes de cours* .

Picchini, U., Ditlevsen, S., De Gaetano, A. and Lansky, P. (2008). Parameters of the diffusion leaky integrate-and-fire neuronal model for a slowly fluctuating signal. *Neural Computation* **20**, 2696–2714.

Pinheiro, J. and Bates, D. (2000). *Mixed-effect models in S and Splus.* Springer-Verlag, New York.

Revuz, D. and Yor, M. (1999). *Continuous martingales and Brownian motion*, vol. 293. Springer-Verlag, Berlin.

Schmisser, E. (2013). Penalized nonparametric drift estimation for a multidimensional diffusion process. *Statistics* **47**, 61–84.

Tsybakov, A. (2009). *Introduction to nonparametric estimation.* Springer, New York.